



Improved premises for cell factory development

An enhanced understanding of established microbial protein production systems

Søgaard, Karina Marie

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Søgaard, K. M. (2015). *Improved premises for cell factory development: An enhanced understanding of established microbial protein production systems*. Novo Nordisk Foundation Center for Biosustainability.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Improved premises for cell factory development

*An enhanced understanding
of established microbial protein
production systems*

Karina Marie Søgaard

Ph.D. Thesis
November 2015



Improved premises for cell factory development

*An enhanced understanding of established
microbial protein production systems*

Ph.D. Thesis

Karina Marie Søgaard

The Novo Nordisk Foundation Center for Biosustainability

The Technical University of Denmark

November 2015

Supervisor: Morten H. H. Nørholm

The work presented in this Ph.D. thesis was performed between December 2012 and November 2015 at the Novo Nordisk Foundation Center for Biosustainability at the Technical University of Denmark and included a research stay at DNA2.0 in Menlo Park, California. The work was supervised by senior researcher Morten H. H. Nørholm from the Technical University of Denmark. Funding was provided by the Novo Nordisk Foundation A/S.

Abstract

The sustainable manufacturing of medicines, materials and chemicals is enabled with biotechnology, and the key to the development of new processes, as well as improvement of existing ones, lies in our fundamental understanding of the biological systems we manipulate. Recombinant protein production is at the core of biotechnology and numerous molecular tools and bacterial strains have been developed over the past four decades for this purpose. Understanding of the genetic code and our ability to manipulate genetic material, paves the way for the microbial cell factory development that enables production of protein in a sustainable, cost-efficient manner. In this thesis I report the joined efforts of my colleagues and myself, to improve the premises for cell factory development by optimizing the cloning strategies, improving the awareness of unforeseen side-effects in complex bacterial expression systems, and building a platform for enhanced expression of certain plant genes in bacteria. The relevance of the conducted research to the field of biotechnology is covered, as well as necessary scientific background and history. Specifically, the surprisingly minor effects of tRNA overexpression on the production of a large number of membrane proteins in *Escherichia coli* are reported, and also the subsequent work elucidating two types of side effects: in some cases growth and gene expression are directly impaired by the extra tRNA sequences and in other cases the apparent positive effects are instead caused by a differential expression of the lysozyme gene encoded on the same plasmid. These phenomena seem to have been largely overlooked despite the huge popularity of the T7/pet-based systems for bacterial protein production. Additionally, the paradox that standardization comes at the cost of reduced flexibility is addressed. The development of a cloning strategy is covered, aiming at maximized standardization while maintaining high flexibility in choice of cloning technique. Finally, the systematic optimization of production of six plant-derived cytochrome P450 enzymes in *E. coli*, using a high-throughput workflow based on C-terminal fusions to the green fluorescent protein is also presented. Our work suggests that there are no inherent limitations in using different standard *E. coli* strains and expression conditions for exploiting cytochrome P450s for biotechnological applications. The results of this thesis have improved the premises for cell factory development in the future.

Dansk resumé

Den grundlæggende forståelse af biologiske systemer bærer nøglen til en bæredygtig produktion af medicin, kemikalier og materialer, og bioteknologien muliggør dette. Produktion af rekombinante proteiner er helt centralt for den bioteknologiske industri. Talrige molekylære værktøjer og mikroorganismer er blevet udviklet gennem de seneste fire årtier, for at udvikle nye processer og forbedre eksisterende, med henblik på netop rekombinant proteinproduktion. Vejen til at udvikle mikrobielle cellefabrikker går gennem en forståelse af den genetiske kode og vores evne til at manipulere arvmasse: Dette gør os i stand til at skabe bæredygtige og omkostningseffektive systemer. I denne afhandling præsenteres mine kollegers og min fælles indsats for at forbedre præmisserne for udviklingen af cellefabrikker gennem i) øget opmærksomhed på uforudsete effekter i komplekse bakterielle produktionssystemer ii) optimering af kloningsstrategier og iii) udviklingen af en platform til ekspression af en bestemt klasse af plantegener i bakterier. Der redegøres for forskningens relevans for bioteknologi, suppleres med nødvendig baggrundsinformation og et historisk perspektiv. Konkret vises det at øget ekspression af tRNA-gener, der afkoder sjældent forekommende kodons har overraskende lille effekt på membranproteiner i *Escherichia coli*, og efterfølgende belyses to typer af sideeffekter ved denne fremgangsmåde: I visse tilfælde er vækst og proteinproduktion nedsat som en direkte effekt af de ekstra tRNA-gener. I andre tilfælde ses det at den tilsyneladende positive effekt på proteinproduktion ikke skyldes tRNA-generne, men forskellige niveauer af enzymet T7 lysozym, som er udtrykt fra samme plasmid. Disse fænomener har været overset, på trods af den store popularitet som T7/pET-baserede systemer til proteinproduktion i bakterier har fået. Ydermere adresseres det paradoks, at standardisering af kloning oftest betyder reduceret fleksibilitet. Udviklingen af en kloningsstrategi sigtet mod maksimal standardisering, hvor fleksibiliteten bevares rapporteres. Endelig gennemgås den systematiske optimering af en bakteriel produktionsplatform. Produktionen af seks cytochrom P450 enzymer fra planter optimeres ved brugen af reporter-proteinet GFP (green fluorescent protein). Dette studie antyder, at der ikke er nogen begrænsning for valg af *Escherichia coli* stammer til produktion af denne type enzymer. Sammenlagt har resultaterne af denne afhandling forbedret præmisserne for udviklingen af cellefabrikker fremover.

Acknowledgements

The scientific journey I have been on since 2012 has been joyful, and I feel been fortunate that I got the chance to pursue my in the best research environment I could imagine. In my view, the contribution of a Ph.D.-thesis is a tiny expansion of existing knowledge, and this is my contribution.

I would like to express my sincere appreciation and gratitude to my supervisor Morten Nørholm. He had barely established his research group when I was lucky to get under his wings, and he jumped into the demanding and maybe sometimes unrewarding task of educating researchers for the future, starting with me. He believed in me, and his always supporting and optimistic perspective have been most welcome.

Furthermore want to take this opportunity to thank all present and previous members of the membrane synthetic biology research group at CfB: Ulla, Morten T, Susanna, Mafalda, Tonja, Sofie, Dario, Se Hyeuk, Virginia, Roberto, Maja, Ida, Emil, Victor, Kristian. I have always been able to count on you for advice, cheerful comments, a helping hand and your dark and geeky sense of humor. I cannot imagine a better group of people to be surrounded with, while facing the ups and downs of research. Several of you helped me read through the final thesis, and with the help of Tonja, the final experiments were conducted in the very last days. I cannot thank you enough.

I also want to thank the people who I came in contact with and who helped me during my Ph.D. studies. Thanks to all the people at the DTU Novo Nordisk Foundation Center for Biosustainability, for making my time pleasant and memorable. Anna Koza and Emre Özdemir supported me with their expertise, and made it possible to add a new dimension to my results. I would like to thank Dan Daley from Stockholm University for providing the membrane protein gene GFP library used for the research presented in Chapter 2.

I was privileged to visit DNA2.0 for an external research stay, and I truly enjoyed my time there. I would like to thank Claes Gustafsson and Mark Welsch for their supervision during this time, and valuable scientific discussions that helped shape the further work. And Koy deserves a special thank for showing me Californian life

outside the laboratory. I also want to thank OticonFonden from whom I received a scholarship to cover some of my expenses for this stay.

I got the chance to join Jesper Berg in representing DTU on a career fair in my first year as a PhD-student, and several ever since. These experiences grounded a seed, and showed me how fun and challenging it can be to interact and communicate with people on many levels, and this has been very valuable to me.

Finally, I would like to mention and thank to my family and friends who have endured the past three years with me and shared their support, and especially my husband, Brian who is always there for me. His unconditional love, positive mood, and deep desire to *actually* understand my work, is invaluable.

List of publications

The results presented in this thesis have formed the basis for or contributed to the following articles and manuscripts:

Different side effects of extra tRNA supply in standard bacterial protein production scenarios

Karina Marie Søgaard, Tonja Wolff and Morten H. H. Nørholm.

(Manuscript planned for submission)

Assembly of highly standardized gene fragments for high-level production of protoporphyrins in *E. coli*

Morten T. Nielsen, Karina M. Madsen, Susanna Seppälä, Ulla Christensen, Lone Riisberg, Scott J. Harrison, Birger Lindberg Møller and Morten H. H. Nørholm.

ACS Synth. Biol., **2015**, 4 (9), pp 959–965 DOI: 10.1021/acssynbio.5b00033

De-bugging and maximizing plant cytochrome P450 production in *Escherichia coli* with a scalable GFP-based optimization scheme

Ulla Christensen, Dario V. Albacete, Tonja Wolff, Morten T. Nielsen, Scott James Harrison, Anders Holmgaard Hansen, Birger Lindberg Møller, Susanna Seppälä and Morten H. H. Nørholm.

(Manuscript planned for submission)

Abbreviations and nomenclature

IPTG	isopropyl β -D-1-thiogalactosepyranoside
T7 RNAP	T7 RNA polymerase
DNA	deoxyribosenucleic acid
A, U, T, C, G	Adenine, uracil, thymine, cytosine, guanine
RNA	ribosenucleic acid
tRNA	transfer-ribosenucleic acid
mRNA	messenger-ribosenucleic acid
CAI	Codon Adaptation Index
GFP	Green fluorescent protein
aaRS	Amino acyl tRNA syntethase
RNase P	ribonuclease protein
P450s	cytochrome P450 monooxygenases
Rosetta	used about the strain RosettaII(DE3)pLysS
pRARE	used about the pRARE2pLysS plasmid

Comments

This thesis is composed of an introductory chapter followed by 3 chapters each comprising a manuscript, either published or prepared for submission, and they are written in a format intended for a scientific journal. The majority of my time has been spent on the research presented in chapter 2, and this is reflected in the introduction, as subjects particularly relevant for this chapter are emphasized. The background information includes *in silico* analysis performed along with the research, and is essential for the conclusions drawn in the research paper.

Table of contents

Abstract	ii
Acknowledgements	iv
List of publications	vi
Abbreviations and nomenclature.....	vii
Comments.....	viii

CHAPTER 1 INTRODUCTION AND BACKGROUND..... 1

1.1 INTRODUCTION AND SYNOPSIS	2
1.2 HISTORY OF BIOTECHNOLOGY IN BRIEF	3
1.3 CELL FACTORIES	3
1.4 THE <i>E. COLI</i> STRAIN BL21: "THE GOLD STANDARD"	6
1.5 POPULAR <i>E. COLI</i> EXPRESSION SYSTEMS	7
1.6 THE T7 EXPRESSION SYSTEM	10
1.7 CODON BIAS AND RELATED CHALLENGES IN HETEROLOGOUS GENE EXPRESSION	12
1.8 STRUCTURE, SYNTHESIS AND FUNCTION OF tRNAs.....	14
1.8.1 <i>tRNA synthesis</i>	15
1.8.2 <i>tRNA function</i>	16
1.9 CORRELATION BETWEEN tRNA CONCENTRATION AND CODON USAGE.....	18
1.10 THE pRARE PLASMID.....	20
1.11 HISTORY OF tRNA-SUPPLYING PLASMIDS	24
1.12 STANDARDIZED CLONING STRATEGIES	27
1.13 HIGH-VALUE COMPOUND PRODUCTION IN <i>E. COLI</i>	28
1.14 FUTURE PERSPECTIVES ON CELL FACTORIES	30
1.15 REFERENCES.....	31

CHAPTER 2 DIFFERENT SIDE EFFECTS OF EXTRA tRNA SUPPLY IN TYPICAL BACTERIAL PROTEIN PRODUCTION SCENARIOS..... 39

2.1 INTRODUCTION	40
2.2 RESULTS.....	42
2.3 DISCUSSION	48
2.4 MATERIALS AND METHODS	48
2.5 REFERENCES.....	52

CHAPTER 3 ASSEMBLY OF HIGHLY STANDARDIZED GENE FRAGMENTS FOR HIGH-LEVEL PRODUCTION OF PROTOPORPHYRINS IN *E. COLI* 56

3.1 INTRODUCTION	57
------------------------	----

3.2	RESULTS AND DISCUSSION.....	59
3.3	CONCLUDING REMARKS	69
3.4	METHODS.....	70
3.5	REFERENCES.....	75
CHAPTER 4 DE-BUGGING AND MAXIMIZING PLANT CYTOCHROME P450 PRODUCTION IN ESCHERICHIA COLI WITH A SCALABLE GFP-BASED OPTIMIZATION SCHEME		80
4.1	INTRODUCTION	81
4.2	MATERIALS AND METHODS	83
4.3	RESULTS.....	88
4.4	DISCUSSION	95
4.5	CONCLUSION.....	97
4.6	REFERENCES.....	98
4.7	SUPPLEMENTARY FIGURES.....	104
CHAPTER 5 CONCLUDING REMARKS		113
APPENDIX.....		114

Chapter 1

Introduction and background

1.1 Introduction and synopsis

The capacity to endure. That is the definition of sustainability. In biosustainability the term is interpreted as the potential of maintaining or increasing the quality of life on earth, without depriving the planet of its resources. The accomplishment of this has become a heated subject of public debate. The world's leaders meet to determine long-term climate goals to preserve our planet for future generations, and The United Nations have increased its focus on sustainability. By September 2015, the Sustainability Development Goals¹ were a reality. The 17 goals have a broad focus with ambitious targets, covering poverty and equality goals, and climate goals to preserve the planet. Researchers in academia and industry are developing sustainable solutions to replace exhaustible resources by renewable supplies. A widespread production of medicine, chemicals and materials from sustainable sources mark the accomplishments of this industry^{2,3}, and show a future where biotechnology offers some of the solutions to reach a sustainable development.

Synopsis

In the constant chase for better, cheaper or easier ways of producing proteins, the multitude of available systems is a major benefit, but the vast amount of options can turn into a downside. There are many ways to reach an optimal solution, and the scientific work described in chapters 2-4 reveals some of the facets of working with microbial cell factory development. A curiosity about an expression system encompassing overexpression of tRNAs, developed to provide universal expression of eukaryotic genes in *E. coli*, founded the work described in chapter 2. We tested the system in a high-throughput manner, and subsequently in small scale, to discover surprising effects of the system, which are important for heterologous gene expression. A desire for a standardized cloning platform without the drawbacks of such already existing technologies fueled the work described in chapter 3, where we successfully obtain state-of-the art levels of our test compound protoporphyrin IX, in two simple design-build-test cycles. Finally, the need for better expression of membrane bound protein genes, particularly plant derived cytochrome P450s, paved the way for our development of an optimal production platform. The work is described in chapter 4, and this contribution thus adds this plant multi-membered key enzyme family to the toolbox for bacterial cell factory design.

1.2 History of biotechnology in brief

Since ancient times biotechnology has been used by humans to manipulate biological systems. The production of beer, bread and cheese dates back to the beginning of documented history. Plants and animals have been domesticated and manipulated without ever cloning a gene in the laboratory. The emergence of classical molecular biology traces back 60 years, when James Watson, Francis Crick and Rosalin Franklin elucidated the molecular structure of nucleic acids⁴. Recombinant DNA techniques were developed in the early 1970s⁵⁻⁸, and by 1981 the first major products hit the market⁹. Today, the industry is a 300 billion dollar market, with growth rates around ten percent through the past five years¹⁰. Production of recombinant proteins has opened a new era, and proteins that were previously difficult to produce are now easily obtained in cell factories developed for the purpose¹¹.

1.3 Cell factories

Life is built on the same basic principles, regardless of life form, and this is the key that made it possible to develop cell factories. The term reveals the concept: a cell factory is a cell manipulated to work like a factory. Depending on the manipulations, the factory can produce chemicals, proteins, enzymes, or maybe the cell itself is the product. The demand for cell factories and technologies that allow for their manipulation is increasing. These technologies are developing at a fast pace and the performance of cell factories is promising for future applications. The choice of a cell factory takes into account several factors with robustness, available engineering tools, and survival under desired process conditions at the top of the list. The Gram-negative bacterium *Escherichia coli* is comprehensively investigated and widely used in industry¹².

The first commercial enzyme was used as a detergent, and reached the market in 1914¹³. When microbial proteases were introduced in washing powders in 1959, the real breakthrough for industrial enzymes took place. In 1965, the major detergent companies started using the first commercial protease from *Bacillus sp.*¹⁴ which was produced by the Danish company Novo (now Novozymes). This was the start for the big business of enzymes for the world known company.

Chapter 1

The full establishment of the cell factory concept happened when the first recombinant pharmaceutical reached the market in the early 80's. The product was human insulin, and this was produced in *E. coli* cell¹⁵. The Food and Drug Administration (FDA) and European Medicines Agency (EMA) have licensed more than 150 protein-based recombinant pharmaceuticals. Out of these almost one third are obtained in *E. coli*. The engineering of strains has allowed for a progressive development of new cell factories with enhanced performance. Early cell factories were a product of untargeted mutagenic strategies and phenotypic selection, but directed genetic and metabolic engineering soon became important players as well¹⁶. In parallel with a strong focus on the well-studied unicellular organisms such as *E. coli* and *S. cerevisiae* as cell factories, high quality proteins have been produced in mammalian cells and insects, where the multicellular complexity can be a strong advantage, as they can perform posttranslational modification which can otherwise be the limiting factor. Microbial species with unusual traits have gained focus as well, including algae, fungi, and moss¹⁶.

Whether aiming at producing a protein, enzyme or chemical, there are several factors to consider. The process of developing a cell factory is often initiated by the choice of host. There are a wide range of hosts to consider, depending on desired traits¹⁷. With the choice of host, the technology available for the project is defined by the availability of molecular tools, equipment and previous research¹⁸. For products requiring post-translational modifications, a eukaryote host can be preferable¹⁹, but often, bacterial hosts are preferred due to their relatively small size, rapid reproduction, and metabolic diversity. Here, a minor selection of bacterial cell factories is covered, and further review of bacterial cell factories is covered by Demain *et al*¹⁷, Rosano *et al*¹⁸, Ferrer-Miralles *et al*²⁰, Westers *et al*²¹ and Morello *et al*²² where as eukaryotic cell factories are reviewed by Demain *et al*¹⁷ and Punt *et al*²³.

Recombinant protein production is among others carried out in the Gram positive strains *Bacillus subtilis* and *Lactococcus lactis*, from which secretion of proteins of interest into the medium is easily obtained, due to their secretory pathways²². This enables easier downstream processing, as the product can be purified directly from the medium²⁰. Endotoxins can be a problem with some bacterial hosts, when producing therapeutics for use in humans, as they are known to induce fever²⁴. *B. subtilis* is free of endotoxins, and this feature is convenient for therapeutics

production, since it eliminates the need for downstream endotoxin removal²². Both *B. subtilis* and *L. lactis* are well studied organisms, and their genomes have been sequenced^{25,26}. Additionally, genetic tools are available, and the behavior under different growth conditions is well documented, making the link from laboratory development to industrial production shorter^{27–30,31}.

Some of the challenges in heterologous gene expression require special traits in a production host. Rather than attempting to overcome these with a traditional host of choice, alternative sources can be considered. For example *Pseudomonas putida* which has extraordinary metabolic versatility genus with a diverse biotechnological potential, as it is highly tolerant to solvents and high temperature³². This enables the production of several bulk and fine chemicals in high titers³³. *P. putida* is a new player in the field of microbial cell factories, and some of the limitations are caused by the early progress-stage, where lack of knowledge in genotype/phenotype relationships poses a constraint in the applications³².

E. coli is a popular choice for recombinant protein production and expression of pathways. Additionally, it is widely used for cloning, genetic modifications, and small-scale production for research purposes. Historically, the bacterium has been widely studied and the development of molecular genetics has mainly been based on *E. coli*³⁴. Likewise, it was the first organism to get FDA-approval recombinant protein production²⁰. Since the organism is well established in each laboratory, it often becomes the first choice of expression host. The advantages of using *E. coli* includes the fast doubling time (20 minutes under optimal conditions³⁵), easily obtainable high cell densities³⁶, and easy transformation with exogenous DNA as well as plasmid purification³⁷.

Many of the challenges encountered when producing a recombinant protein in *E. coli* are general to cell factories, but often context dependent. For example, product toxicity can have fatal consequences for the host. This issue can be resolved by controlling the expression with a different, tighter promoter system, a weaker promoter, or lower induction concentration. Ultimately the solution can be a different choice of host. The formation of inclusion bodies, intracellular granules of protein³⁸, poses a major bottleneck in recombinant protein production³⁹. In inclusion bodies, the protein is inactive and cannot perform its task. When inclusion bodies are a persistent problem, several strategies exist to downstream processing of the proteins, to obtain a

fully functional product⁴⁰. The therapeutic use of recombinant proteins from *E. coli* is also complicated by the accumulation of endotoxins as mentioned earlier. It requires an additional purification step to have a final endotoxin-free product²⁴. Below follows a more detailed overview of the expression system often applied when expressing heterologous genes in *E. coli*, and section 1.4 contains an elaboration on *E. coli* cell factories.

1.4 The *E. coli* strain BL21: "the gold standard"

Despite the worldwide use of *E. coli* for experimental biology, the majority of *E. coli* laboratory strains descend from a few individual isolates⁴¹: K-12, often used for cloning, and the B-strains, often used for expression. K-12 has a straightforward history: It was isolated from a patient suffering from convalescent diphtheria in 1922⁴². The B-strain has a more complex origin, being shared and renamed between researchers for decades. The strain has been traced back to 1918, when it was used at the Pasteur Institute in Paris⁴¹. When fast-forwarding to the present, numerous derivatives of both strain types exist. Derivatives of K-12 such as DH5 α and DH10b are often used for cloning, and they are especially suitable for this purpose due to their recombination and endonuclease deficiencies, and the possibility to perform blue/white screening (not covered here, but described in detail by H. Miller⁵³ and S. T. Smale⁵⁴).

In *E. coli* laboratories one strain in particular has become the gold standard for recombinant protein expression: BL21(DE3). It has a genome integrated T7 RNAP, and together with vectors where T7 promoters controls the gene of interest, it comprises the T7 expression system^{45,46}. This system was developed and patented^{45,46} by William Studier and his colleagues⁴⁵, and is described in detail in section 1.6. This innovation is still highly relevant as more papers are being published every year, as depicted in Figure 1.

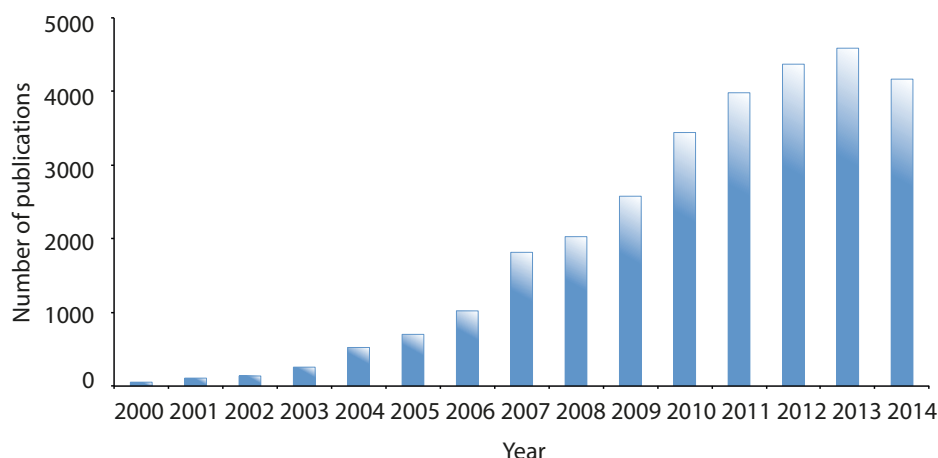


Figure 1 Number of search hits pr. year for "BL21(DE3)" at pubmed.org

BL21 and derivative strains hereof are advantageous, as they are deficient in Lon and OmpT proteases. OmpT is known to degrade T7 RNAP, and *lon* mutants are known to exhibit decreased protein degradation⁴⁷. Since the introduction of the strain in 1990 several derivative strains with traits specific for certain applications have been developed⁴⁸. Several derivatives supply the genes for a number of minor tRNA species, to overcome a codon bias (covered in section 1.7), and one of the most remarkable variants is the strain Rosetta2(DE3)pLysS (hereafter referred to as Rosetta).

The work presented in chapter 2 of this thesis is focused on the applications and limitations within the Rosetta strain. Many of the additional variations of BL21 are reviewed by Kay Terpe⁴⁸. The Rosetta contains the T7 expression system and a pRARE plasmid, which are both described in the following sections. The underlying assumptions for the design of this particular strain have its roots in the details of variability in the genetic code between species. Prior to a thorough review of the pRARE plasmid harbored by the Rosetta strain, the genetic principles are covered.

1.5 Popular *E. coli* expression systems

In laboratories all over the world, *E. coli* is the preferred choice of host cell, and the positive consequence of this is a broad collection of tools for protein expression in *E. coli*. The most commonly used expression systems are listed in Table

1. Each system has certain key features making them suitable for different types of protein expression.

Table 1 Popular *E. coli* expression systems and their key features. Adapted from K. Terpe⁴⁸.

Promoter	Induction method	Key features
P_{lac} ⁴⁹	IPTG	Relative low-level expression; titratable, leaky
P_{trc} and P_{tac} ⁵⁰	IPTG	High-level expression; titratable, leaky
P_{T7} ⁴⁵	IPTG/rhamnose	Very high-level expression, titratable, leaky
Phage λ P_R , P_L ⁵¹	Temperature shift	High-level expression; tight control
P_{tetA} ⁵²	(Tetracycline) or anhydrotetracycline	Mid- to high-level expression; tight
P_{araBAD} ⁵³	L-arabinose	Low- to high-level expression; titratable; tight
P_{rha} ⁵⁴	L-rhamnose	Low- to high-level expression; titratable; tight

The expression systems P_{lac} ⁴⁹, P_{trc} ⁵⁰ and P_{tac} ⁵⁰ are all induced with the lactose-analog isopropyl β -D-1-thiogalactopyranoside (IPTG). This compound is used for induction of P_{T7} ⁴⁵ as well, but this action can also be performed with rhamnose, depending on strain system. Two different inducers can activate the same system, as induction is not directly of the promoter. The T7 promoters are not recognized by *E. coli*'s own RNA polymerase, and the system is thus dependent on its own phage-derived T7 RNA polymerase (T7 RNAP)⁴⁵. Induction activates transcription of the polymerase from a P_{rha} or P_{lacUV5} promoter, subsequently allowing transcription from T7 promoters. IPTG is an expensive inducer, and because of this these systems are most applicable for production of high-value compounds. Due to the low intracellular levels of protein obtained with P_{lac} , this promoter is especially suitable for membrane protein production, or production of gene products which are toxic to the cell⁴⁹. All

Chapter 1

three promoter-systems P_{lac} , P_{trc} and P_{tac} are regulated by catabolite repression, and the T7 expression system was developed to offer an expression system free from catabolite repression. The IPTG-induced T7 expression system is an important element of the experimental work described chapter 2 and 4, and is thus further portrayed in section 1.6.

The λ phage promoters P_L and P_R also exploit the specificity of the phage genes, but in a different manner. Here, the promoter is controlled by the cI-repressor. The system is neat in its functionality since below 29°C the repressor is inactive, leaving the promoter available for constitutive expression. From 29°C and above, expression is repressed and then released again above 42°C⁵¹.

The tetracycline inducible P_{tetA} system is a popular choice for heterologous gene expression due to its tight nature. The system relies on the repressor TetR, which is not encoded by an *E. coli* gene. The repressor is supplied on a plasmid, so the system is independent of strain choice. The induction is performed with an antibiotic which is clearly problematic, however, the use of anhydrotetracycline rather than tetracyclin relieves growth interference problems, and gives 35-fold higher promoter binding with a 100-fold less active antibiotic⁵⁵.

The sugar-induced promoters P_{araBAD} and P_{rha} are useful alternatives when a tight and titratable system is required for protein expression. The AraC activator controls expression from P_{araBAD} , and a high level of arabinose in the growth medium yields high expression. In absence of arabinose, the expression is tightly shut down⁵³. The mechanism for gene synthesis from P_{rha} relies on induction performed with L-rhamnose, which activates transcription of the genes *rhaS* and *rhaR*. These genes regulate the L-rhamnose catabolism by activation of the rhamnose uptake and metabolism⁵⁴, which allows transcription from P_{rha} .

Due to the low cost, high productivity and rapid use, bacterial systems and especially *E. coli*, remain the most attractive choice for recombinant protein production⁵⁶. The rational choice of expression system can determine failure or success, and it is thus important to understand the characteristics of the desired product, to enable informed decision-making.

1.6 The T7 expression system

The *E. coli* T7 expression system is very popular among researchers due to the specificity and potential for high intracellular target protein production⁴⁵. The system is based on an *E. coli* lysogen of DE3, a phage lambda derivative carrying the T7 "1" gene, which encodes T7 RNAP controlled by the lacUV5 promoter⁴⁵. This promoter is a strong variant of the wild type lac promoter, and being insensitive to catabolite repression, it is only controlled by the lac repressor lacI, which binds to the lac operator⁵⁷.

The T7 expression system is shown in Figure 2. Bacteriophage T7 RNAP is highly selective for its own promoters which do not occur naturally in *E. coli*⁵⁸, and is additionally very efficient in initiating transcription, and the elongating rate is around 5 times faster than for *E. coli* RNA polymerase^{58,59}. The production of T7 RNAP requires induction, and once the T7 expression system is active, the target protein produced from T7 promoters can constitute more than 50% of total cell protein in a few hours⁴⁵.

An important addition to the system was soon developed, to prevent leaky expression of T7 RNAP: T7 lysozyme, known to inhibit T7 RNAP, expressed on a plasmid⁶⁰. Accumulation of T7 lysozyme does not prevent growth of *E. coli* cells despite of the enzyme's amidase activity. The lysozyme cannot penetrate the inner membrane, and is therefor prevented from reaching its substrate, the peptidoglycan layer of the outer membrane⁶⁰. Several T7 lysozyme-expressing plasmids were developed simultaneously for different purposes. They all carry the *cat* gene for chloramphenicol resistance, and the P_{tet} recognized by *E. coli* RNA polymerase. Both plasmids pLysS and pLysE have a $\Phi 3.8$ promoter ($P_{\Phi 3.8}$), which is recognized by T7 RNAP, downstream of the gene.

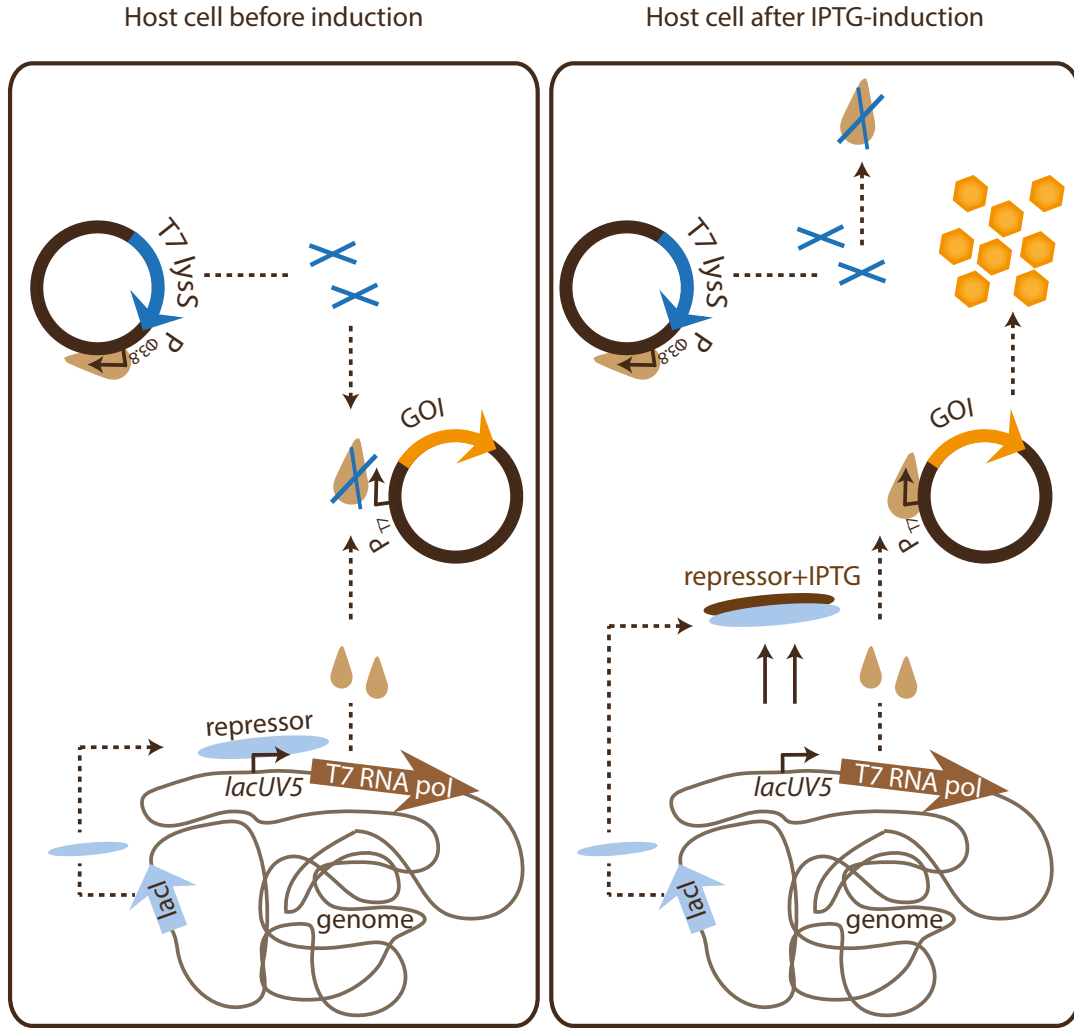


Figure 2 The *lac* repressor LacI binds to P_{lacUV5} and inhibits transcription of T7 RNAP. The repression is not tight, and leaky expressed T7 RNAP is prevented from transcribing GOI from P_{T7} by T7 lysS. T7 lysS is expressed from a plasmid, driven by T7 $P_{\phi 3.8}$. Induction is performed with IPTG, which binds the *lac* repressor and allows transcription of T7 RNAP. T7 lysS is heavily diluted by T7 RNAP, and T7 RNAP can subsequently transcribe GOI from P_{T7} .

This allows the T7 RNAP to transcribe around the plasmid, and produce small amounts of T7 lysozyme mRNA^{46,61,100}. Additionally, the lysozyme gene is oriented differently on the two plasmids, so pLysE allows for production of T7 lysozyme from P_{tet} as well, whereas pLysS only allows lysozyme to be produced from $P_{\phi 3.8}$. Two other plasmid variants, variants, pLysL and pLysH have no $P_{\phi 3.8}$, and can only produce lysozyme from P_{tet} . An additional variant of the lysozyme, LysY⁶² exists, which lacks the amidase activity, and can thus not perform cell lysis. This has been used to develop the pLEMO21 expression system, where P_{rha} allows for titration and controls expression of the T7 lysozyme gene⁶³. In the work presented chapter 2, the

pLysS version of the T7 lysozyme plasmid is applied in combination with either BL21(DE3) or the tRNA-supplying Rosetta(DE3), and the following section will unravel the biological importance of tRNAs.

1.7 Codon bias and related challenges in heterologous gene expression

Upon deciphering of the genetic code the degeneracy became apparent⁴. Between one and six different nucleotide codons can encode a single amino acid, which allows for synonymous codon choice. The role of synonymous codon choice have puzzled scientists ever since, and it was soon discovered that the frequency for using a certain codon varies between different organisms⁶⁴, a phenomenon termed codon bias. This bias seems restricted to simpler organisms, and have been observed in a broad variety of these^{64,65}. When transferring genetic material from one genome to another, this bias can be problematic as the host might not transcribe and/or translate the foreign genetic sequence in an optimal manner. Partly, this may be caused by a parallel low abundance of transfer-RNAs (tRNAs) cognate to codons that are low in frequency – these codons are commonly referred to as rare codons^{66,67}. Rare codons are defined by being used at a frequency below 0.05 % in the genome, and being decoded by less abundant tRNAs⁶⁸. These tRNAs are interchangeably referred to as rare or minor tRNA species. For *E. coli*, this definition yields the following rare codons: CGA, CGG, AGG, GGA, AGA and AUA.

Numerous studies have shown that synonymous codon mutations can have drastic consequences for gene expression. Factors such as mRNA stability and structure can be affected, but also initiation and elongation on a translational level, as well as protein folding^{69–72}. It is possible that the redundancy in the genetic code has evolved as an extra layer of information stored in DNA – e.g. as a method to preserve structural information of proteins within the nucleotide content⁷³. Early genetic research revealed a positive correlation between expression levels of genes and the codon usage⁷⁴, and several strategies for codon optimization have been developed to optimize heterologous genes for expression in the desired host.

The Codon Adaptation Index (CAI)⁷⁵ is established based on a reference set of highly expressed genes, from which the optimal codon composition for an organism

can be defined. The underlying assumptions are that translational selection has optimized gene sequences for highly expressed genes. These genes presumably compete for resources, and thus require more accurate translation⁷⁵. The reference gene set is used to compute the weight of the codons for each amino acid. The mean value of all codon weight in a sequence defines their CAI-value. This means, that if a gene is composed exclusively of the most frequently occurring codon for each amino acid in the reference set, it will have a CAI-value of 1. Successful expression of genes designed to match host bias or maximizing CAI have been reported in many cases^{76,77}. Nevertheless, a defined relationship between these optimization strategies and expression has not been found⁷⁸. The simplification of codon usage in CAI has been criticized⁷⁹, and a recent study showed that 81 synthetic genes exhibited no correlation between the genes' CAI-values and their actual level of gene expression⁷⁸.

Several prediction models have been developed to estimate the consequences of codon bias and tRNA availability for gene expression. One example is the effective number of codons: a model, which provides a measure of the extent of codon preference in a gene. By comparing the actual codon composition of a gene to a theoretical equal use of synonymous codons, an effective number of codons between 20 and 61 is given⁸⁰. Other examples include the aforementioned CAI⁷⁵, and the tRNA adaptation index⁸¹. The tRNA adaption index is different from the other prediction models, because it considers the interactions between the tRNA and the codon, and the effects of wobble pairing, where other pairs than A-T and G-C are formed⁸² (Wobble base pairs are explained in section 1.8). Wobble pairing is very common for tRNA-codon interactions⁸³, and is thus important to consider for this type of predictions. A recent framework also encounters several protein synthesis mechanisms: It is shown that for some tRNA species, there are very few molecules available for the translating ribosome and additionally that there is a dependency between the codon-specific elongation rate and the overall codon usage in the cell⁸⁴. Recent studies have challenged some of these prediction theories. With a set of 154 synthetic variants of green fluorescent protein (GFP), all encoding the same protein sequence but with varied codon compositions, the widely used prediction theories were put to the test⁶⁹. The gene set contains sequences with both high, intermediate and low frequencies of rare codons. Kudla and coworkers (2009)⁶⁹ showed that neither the CAI nor frequency of optimal codons could predict the fluorescence

levels. On the contrary, some of the highest expressed genes showed low CAI values. The study revealed that the structure of the mRNA is a determining factor for gene expression levels, as weaker structures around the start codon yielded high gene expression⁶⁹. This work was further supported when the hypothesis that rare codons in the 5' end of genes may positively influence translation efficiency⁸⁵ was recently verified⁸⁶. By assaying the expression of 14.000 synthetic reporter genes with codon usage variation in the 5' end, it was shown that the expression of a single gene could be improved up to 14-fold by exchanging N-terminal common codons for rare codons. Using a computational prediction of the genes' mRNA structure, it was shown that a reduction in the RNA structure free energy is responsible for the increased expression, not the fact that the codons are rare. The link between expression levels and rare codons is observed, since they often contain A/T in the third position, which affects the folding energy of the RNA⁸⁶.

It has very recently been shown that the region containing the junction of the DNA and the vector, situated between the Shine-Dalgarno sequence and the start codon, has an impact on gene expression levels as well. By varying the sequence of the junction, high-level expression was achieved for genes previously difficult to overexpress, and the work reveals the importance optimizing or avoiding scarring sequencing when inserting a gene into a vector⁸⁷.

1.8 Structure, synthesis and function of tRNAs

Acting as the physical link between the genetic code and the protein sequence, tRNAs are essential to protein synthesis. One of the characteristics of tRNA molecules is the ability to fold into a tight structure⁸⁸. A tRNA molecule consists of 75-90 nucleotides and the secondary structure of tRNAs resembles a cloverleaf, whereas the tertiary structure often is shown as an L-shape. The cloverleaf structure is depicted in Figure 3. The structure of the molecule enables the tRNA to fit into the ribosome. The three stems in the structure are denoted the D-stem, the

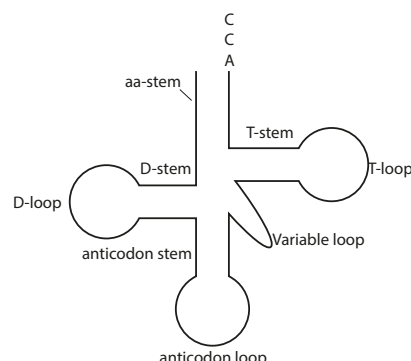


Figure 3 Schematic representation of a classic secondary tRNA structure

anticodon stem and the T-stem, and they each have a loop denoted the D-loop, anticodon loop and T-loop respectively. The T-arm (stem and loop) is the recognition site for the ribosome, where the ribosome-tRNA complex is allowed to form during protein biosynthesis. The D-arm is named after dihydrouracil, a base found in the D-loop. This part of the tRNA is important for ribonuclease P (RNase P) recognition but not for interaction⁸⁸. RNase P is important during tRNA synthesis, which is described in section 1.8. The 3' end of the tRNA molecule carries a CCA-tail necessary for amino acid charging⁸⁸. This sequence is also important for base pairing with RNase P RNA during maturation⁸⁹. The amino acid is covalently attached to the terminal adenosine of the tRNA. *E. coli* has the 3' CCA end encoded in all tRNA genes⁹⁰, but this is highly variable among different bacteria, and the sequence can be post-translationally added as well.

1.8.1 tRNA synthesis

The number of tRNAs is vastly varied in all organisms⁸⁸. Moreover tRNAs are often in abundance, as you need less than 61 to read all codons (due to wobble pairing). Of the 86 tRNAs in *E. coli*, 26 are transcribed as monocistronic transcripts, where the tRNA is transcribed alone with appropriate extensions required for further processing. The remaining 60 tRNAs are organized in polycistronic transcripts with up to 7 tRNAs encoded together⁹¹. These transcripts are subsequently cleaved by ribonuclease P (RNase P)⁹². Some of the tRNAs are encoded in rRNA operons, this is a common feature for bacterial rRNA operons⁹³.

Initiation of transcription happens upstream and terminates downstream of the tRNA gene. Most tRNA species are transcribed with extensions in both the 5' and 3' end of the molecule⁹⁴. RNase P is responsible for the formation of a mature 5' end of the tRNA molecule, and it removes the 5' leader sequence by specific cleavage of the sequence immediate upstream of the mature tRNA sequence⁹². RNase E removes the 3' extension formed upon translation by cleavage downstream of the 3' CCA. This action is usually followed by digestion of the remaining nucleotides, until the 3' CCA mature end of the tRNA is formed⁹⁵.

Amino acyl-tRNA synthetases (aaRS) are responsible for the charging of tRNAs. The side chain of the amino acid is recognized by the aaRS, followed by

binding, a reaction that sometimes induces conformational changes in the aaRS. This increases the probability of correct tRNA binding, thus reducing the frequency of miss-charging. The anticodon of the tRNA is the most important feature used by the aaRS to ensure the tRNA matches the amino acid⁹⁶.

1.8.2 tRNA function

As mentioned, the tRNA molecules provide the link between the genetic code and the protein's amino acid sequence through ribosomal and mRNA interactions⁹⁷. Base pairing with the mRNA strand happens through the codon on the mRNA and the anticodon on the tRNA. Watson-Crick base pairing (A-T- and G-C-pairing) connects the first two nucleotides in the codon to the anticodon, while the third can form a wobble-pair⁸² (Figure 4) Wobble pairing among others depends on the hydrolytic deamination of adenosine to inosine, a common and often essential tRNA modification in the wobble position.⁸⁸ Inosine can base-pair with A, U or C, and allows for tRNAs to pair with other codons than their most obvious partner⁹⁸. There are four main wobble base pairs: G-U, I-U, I-A and I-C.

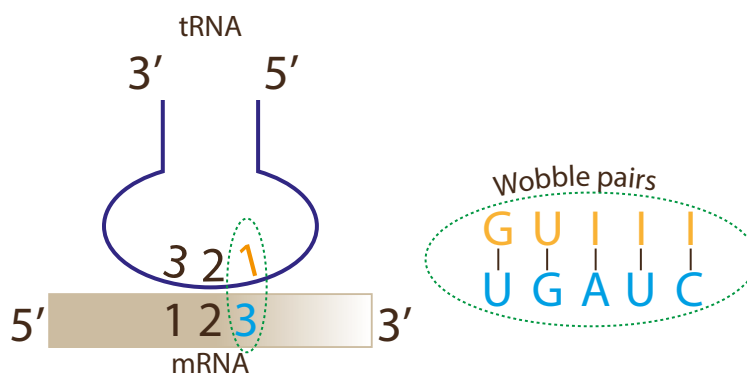


Figure 4 Wobble pairing between tRNA and mRNA. When the first base in the anticodon is a C or G, no wobble pairs can form. For G, U or I in the first position of the anticodon, wobble-pairs G-U, I-A, I-U and I-C can form.

The tRNA carries the encoded amino acid in the other end of the molecule, and is referred to as 'charged'. An aaRS charges the tRNA with its cognate amino acid. The amino acid enters the aaRS together with ATP, followed by AMP being joined to the amino acid upon the release of phosphate. The tRNA takes the AMP's place in the aaRS, which catalyzes the formation of an ester bond between the adenosine residue on the tRNA and the amino acid. The charged amino acid is now released from the aaRS. The elongation factor EF-Tu complexed with GTP is responsible for bringing

Chapter 1

the charged tRNA to the ribosome⁹⁹. Entry of tRNA into the ribosome happens on a trial-error basis: If the tRNA does not match the mRNA codon, the tRNA is rejected. When the correct tRNA enters the ribosome, several conformational changes are performed to accommodate the tRNA¹⁰⁰. The ribosome has three active sites: the A-site, P-site and E-site¹⁰¹. Entry of tRNAs into the ribosome happens in the A-site. The ribosome catalyzes the formation of a new peptide bond, and the peptidyl-tRNA accommodated in the P-site transfers the growing peptide chain to the amino-acyl tRNA in the A-site. This can only happen due to the dramatic conformational change the ribosome undergoes. Upon formation of the peptide bond, the tRNA in the P-site has a free 3' end, and the tRNA in the A-site carries the polypeptide chain. The conformational change of the ribosome brings the residing tRNAs to the next active site. The next elongation cycle can happen after binding of elongation factor G to the ribosome, which brings it back to a conformational state capable of receiving a new tRNA¹⁰¹.

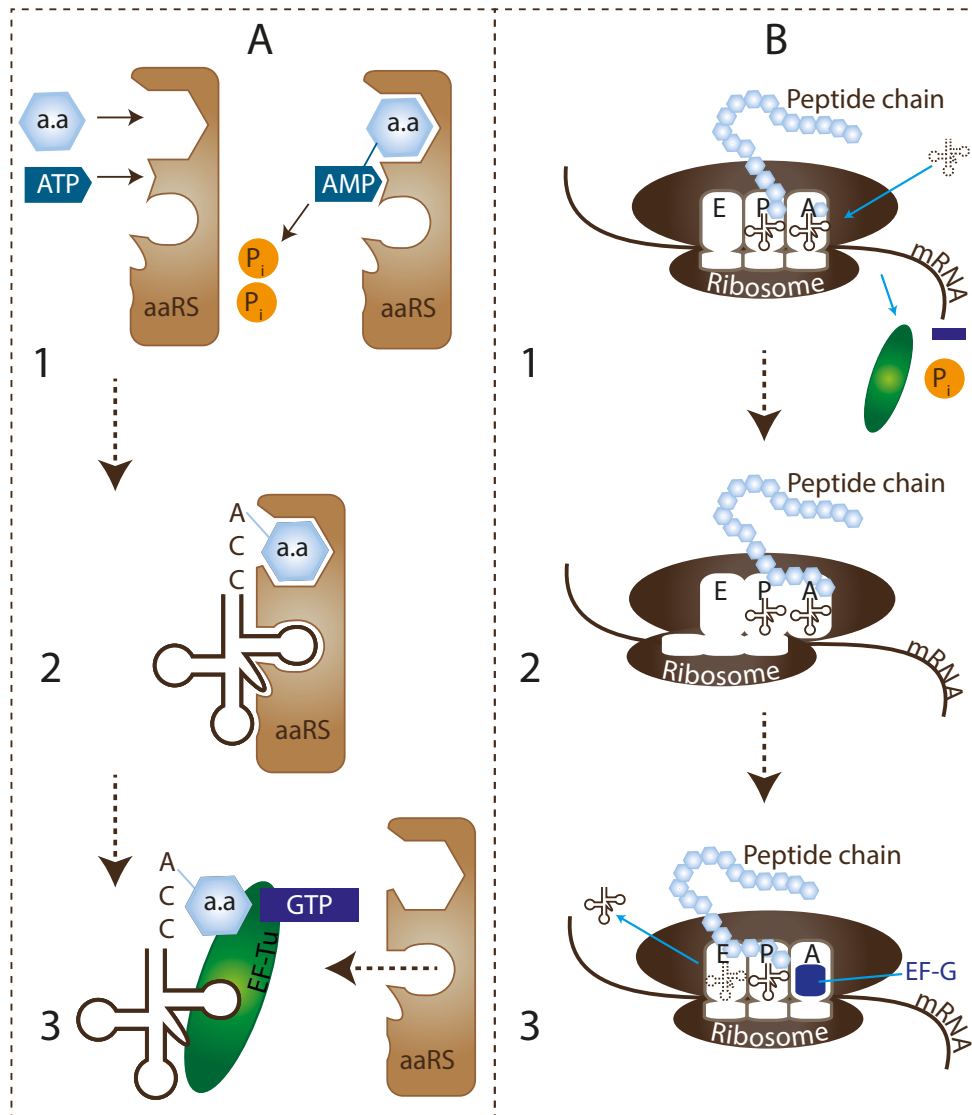


Figure 5. A) Charging and release of tRNA from aaRS. A1) The amino acid enters the active site of aaRS followed by ATP. Upon release and subsequent breakdown of pyrophosphate to phosphate, AMP is bound to the amino acid. **A2)** tRNA replaces AMP in aaRS, and the synthetase catalyzes the formation of an ester bond between the tRNAs adenine-residue and the amino acid. **A3)** The charged tRNA is released from the aaRS and bound by EF-Tu complexed with GTP, which is responsible for bringing the charged tRNA to the ribosome. **B) Interaction of tRNA and ribosome. B1)** The tRNA entered the ribosomes A-site, and EF-Tu, GDP and P_i are released. **B2)** The tRNA in the P-site transfers the peptide chain to the tRNA in the A-site, and the ribosome undergoes dramatic conformational change. **B3)** The tRNA accommodated in the ribosome transfers to the adjacent active site, and EF-G occupies the A-site, to bring the ribosome back to its RNA-accepting state. The tRNA in the E-site leaves the ribosome.

1.9 Correlation between tRNA concentration and codon usage

When the degeneracy of the genetic code became apparent, and understanding of the central dogma was refined, vivid discussion on the correlation between the concentration of tRNAs and the usage of codons took place. It is a complex subject. In BL21, there are 86 tRNA genes as mentioned above. They are used to recognize 61

codons, and charged by 20 amino acyl tRNA synthetases¹⁰². To add even more complexity, many tRNAs can recognize several different codons, due to wobble base pairing⁸².

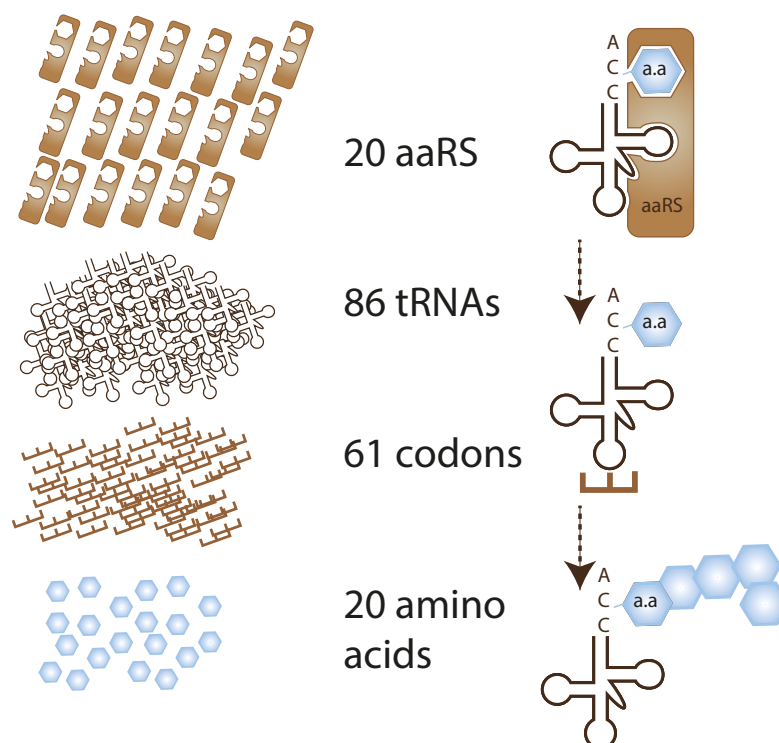


Figure 6 Schematic representation of the complex relationship of aaRS, tRNAs, codons and amino acids. 20 aaRS share the task of charging 86 tRNAs with 20 amino acids, encoded by 61 codons.

Early research suggested that there is a strong relationship between the codon usage and the tRNA pool¹⁰³, and the hypothesis of codon adaptation to the tRNA pool was confirmed by Ikemura⁶⁵, to some extent this work explained the codon bias in highly expressed genes. The understanding of tRNA gene copy number and codon usage evolution became more sophisticated when the co-evolution of tRNA abundance and codon usage was analyzed by Michael Bulmer in 1987¹⁰⁴. With a focus on *E. coli* and yeast, he suggested that rapid division, which causes a strong selection pressure, could partially explain the high codon bias observed in some genomes like *E. coli* and *S. cerevisiae*. The distribution of tRNA abundance has also been shown to vary with the growth rate of *E. coli*, and can be roughly correlated with the bias of the mRNA population¹⁰⁵. The tRNA population in a cell thus closely reflects the codon bias of the mRNA population¹⁰⁵. So when the growth rate of a cell

culture increases, the tRNA species cognate to abundant codons increase in concentration. Most tRNAs cognate to less abundant codons remain constant¹⁰⁵.

The charging of tRNAs is efficient as the charged fraction of all tRNAs are approximately 80% during exponential growth¹⁰⁶. The rate with which the ribosome's A-site is filled closely reflects the concentration of the tRNA able to read the present codon. In the case where an amino acid becomes the limiting factor for protein synthesis, it may happen that the charged fraction of its tRNAs drop dramatically¹⁰⁶.

1.10 The pRARE plasmid

One of the popular *E. coli* strain choices for heterologous gene expression is a commercial strain called Rosetta, available from Merck Millipore (previously Novagen)¹⁰⁷. The strain comes in DE3 and pLysS versions, and is characterized by the tRNA-supplying plasmid called pRARE. This plasmid is a key component of the experimental work presented chapter 2 where the effects of co-expressing this plasmid for increased protein production are investigated, revealing that co-expression of pRARE has very little effect on the production of a large number of membrane proteins in *E. coli*, and that the plasmid can cause impaired growth. We additionally show that other factors than the tRNAs may very well be responsible for positive effects on recombinant protein production.

Since the plasmid is a commercial product, not much is found about it in the literature. With a curiosity about the design of the plasmid, we decided to sequence it to gain a better understanding of the design. The pRARE plasmid is shown in Figure 7 together with the parent plasmid pLysS⁶⁸. The major elements on the two plasmids are the p15A origin giving a plasmid copy number of 10-12¹⁰⁸, the chloramphenicol resistance gene, and the T7 lysozyme encoding gene. The plasmid additionally carries the tRNA sequences for the 7 tRNAs cognate to the most rare *E. coli* codons⁶⁸.

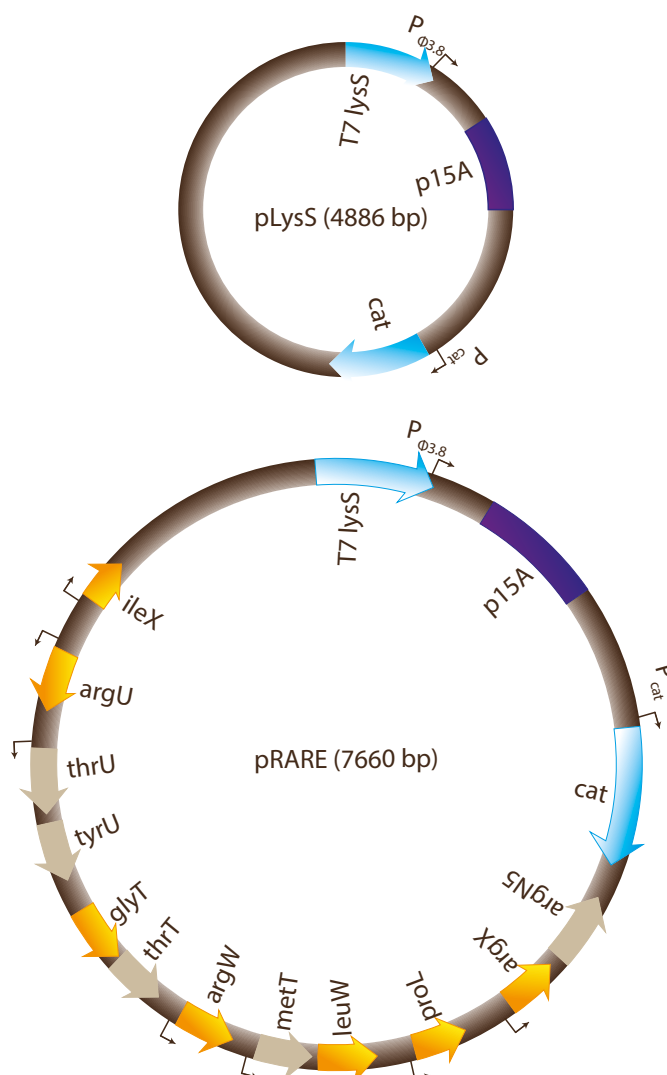


Figure 7 Schematic representation of pLysS and pRARE. Not drawn to scale. Both plasmids contain T7 lysS (blue), p15A (purple) and cat (blue), and pRARE additionally contains 7 rare (yellow) and 5 non-rare tRNAs (beige). The tRNA promoters are placed based on assumptions from the analysis of pRARE.

The *cat* gene encodes chloramphenicol acetyltransferase, which confers resistance to chloramphenicol. This enzyme catalyzes the formation of a covalent bond between an acetyl group from acetyl CoA and chloramphenicol, which disables the binding of chloramphenicol to the 30S subunit of the ribosome. In absence of chloramphenicol acetyltransferase, this bond would inhibit cell growth preventing protein production¹⁰⁹.

pRARE carries a set of tRNAs designed to compensate for minor tRNA species in *E. coli* during heterologous gene expression. They are under the control of their native promoters. Several of these can form wobble pairs, shown in Table 2

Chapter 1

Table 2. List of tRNA genes present on pRARE, their anticodons and respective codons recognized either through Watson-Crick or wobble pairing. Rare codons are shown in bold. **IleX* can pair with AUA due to lysinylation of the C-residue.

tRNA gene	anticodon	codon(s)
<i>argN5</i>	UCG	CGA/CGG
<i>argX</i>	CCG	CGG
<i>proL</i>	GGG	CCC/CCU
<i>leuW</i>	UAG	CUA/CUG
<i>metT</i>	CAU	AUG
<i>argW</i>	CCU	AGG
<i>thrT</i>	GGU	ACC/ACG
<i>glyT</i>	UCC	GGA/GGG
<i>tyrU</i>	GUA	UAC/UAG
<i>thrU</i>	UGU	ACA/ACG
<i>argU</i>	UCU	AGA/AGG
<i>ileX</i>	CAU*	AUA*

It appears that the genes have been directly transferred from the genome of *E. coli* BL21 to this plasmid, as several of them are accompanied by non-minor tRNAs which they are co-transcribed with in the genome. The genomic arrangement of the rare tRNAs are shown in Figure 8, which account for the expected source of 11 of the 12 tRNAs found on the pRARE plasmid. The tRNA gene *argN5* is not accounted for, since it is not found in the genome of BL21. ArgN5 is cognate to the rare codon CGA, and found in the *E. coli* strains O111:H- str. 11128 and O157:H7 str. Sakai, two pathogenic *E. coli* strains¹¹⁰. BL21 does not have a tRNA with the UCG anticodon for decoding of the rare CGA codon - this is translated by pairing with an ICG anticodon¹¹¹.

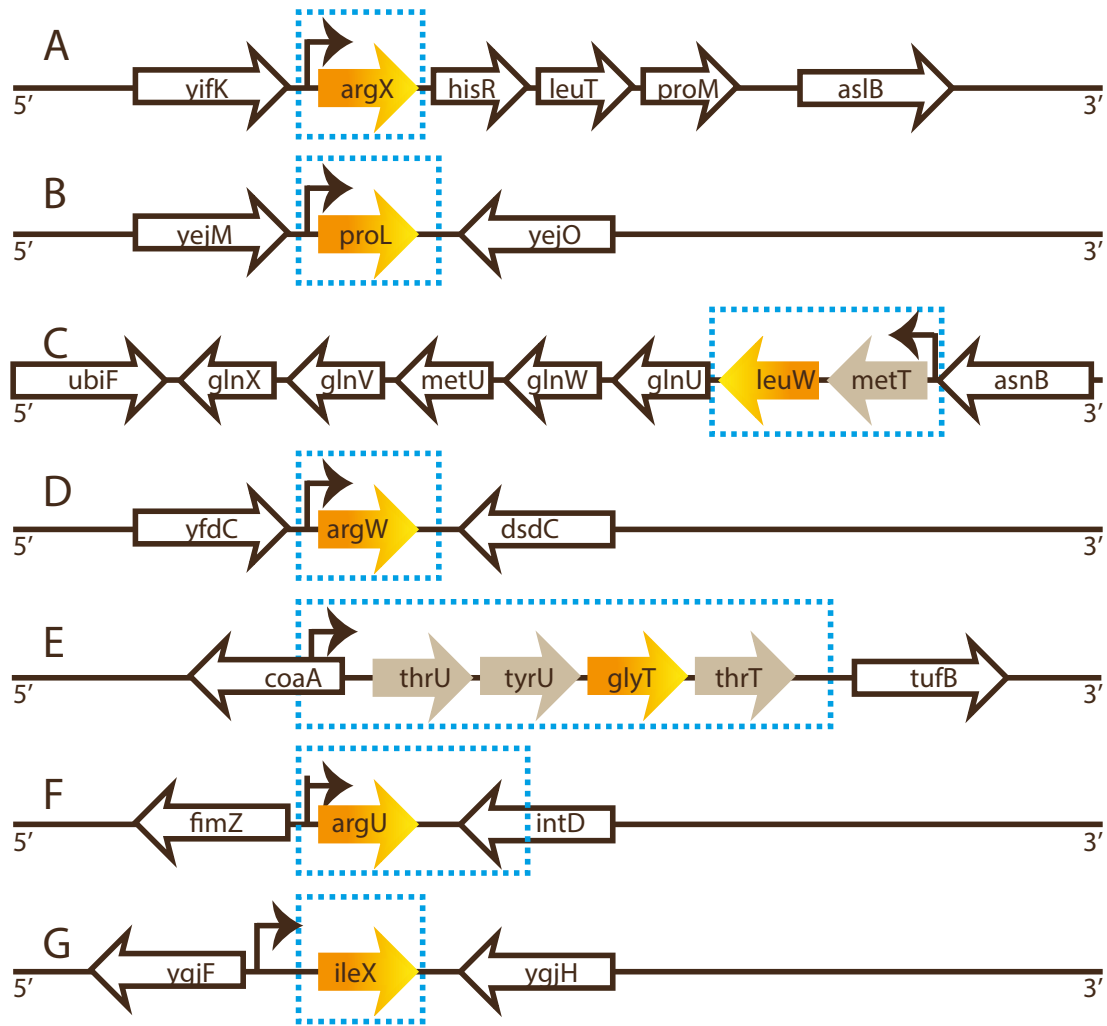


Figure 8 Schematic representation of the genomic arrangement of tRNAs present on pRARE . Blue boxes illustrates the approximate part of the genome that is assumed to have been transferred to pRARE. A) *argX* B) *proL* C) *leuW* D) *argW* E) *glyT* F) *argU* G) *ileX*.

When we comparing the sequence of the pRARE plasmid to the genomic DNA containing the tRNAs, several observations were interesting regarding the design of pRARE. The tRNA genes *argX* and *argW* are found on pRARE with their respective promoter sequences from the genome, and apart from the 5' leader sequence, they are not accompanied by any additional coding sequence. To some extent this goes for *proL* as well, but pRARE additionally contains DNA aligning to the 3' end of the upstream gene *yejM*. In pRARE the tRNA genes *leuW* and *glyT* are both found in operons with other tRNA genes, likely to ensure the transcription from their native promoters. *LeuW* is accompanied by the tRNA gene *metT*, and for *glyT*, the full operon containing 4 tRNA genes is found in pRARE. In the genome, the promoter for this operon is buried in the 5' end of an upstream, reverted gene, and

likewise, the pRARE contains DNA that aligns to this part of the genome. In the pRARE plasmid, *ileX* is encoded in the opposite direction of all other tRNA genes. Curiously, the native promoter does not seem to be present on pRARE. However, it is possible to detect a -10 and -35 region, which most likely serve as a promoter for *ileX*, immediately related to the gene in the plasmid. The base pairing between *ileX* and its cognate codon is special: *ileX* carries the anticodon for methionine (AUG), but the C residue is lysinylated and this reaction modifies the residue to lysidine, a derivative of cytidine¹¹², subsequently allowing AUA base pairing rather than AUG. This is a posttranscriptional modification which is essential for acylation by isoleucyl tRNA synthetase with leucine, as well as the AUA codon recognition¹¹³. The success of overexpressing tRNA genes from this plasmid to achieve high titers of functional protein is different to evaluate to an inherent bias against failed experiments in the literature. In our hands, outcomes of experiments have been difficult to predict, and of high variability. The pRARE plasmid is a complex construction and the rationale behind the design is not available in scientific literature. In the parent plasmid, pLysS, $P_{\phi 3.8}$ is recognized by T7 to transcribe around the plasmid resulting in minor levels of lysozyme. The importance of increased plasmid size for lysozyme levels has never been investigated, and this is discussed in further detail in chapter 2. This chapter also discusses how some of the issues encountered with recombinant protein production in the pRARE in our hands can be explained by some of these details presented here about the plasmid.

1.11 History of tRNA-supplying plasmids

The use of extra copies of minor tRNA species to increase heterologous gene expression started with Brinckmann and co-workers who co-expressed the tRNA gene *dnaY* (now known as *argU*) with eukaryotic genes known to express poorly in *E. coli*¹¹⁴. The genes had a high content of AGA and AGG codons, which are recognized by tRNAAGG/AGA produced from *dnaY*. When supplying *dnaY* on a plasmid, improved accumulation of product up to 30% of total cellular protein was achieved¹¹⁴. The levels of the *dnaY* gene product were shown to have significantly increased when supplying the gene on a plasmid, and inactivation of *dnaY* by deletion resulted in the same phenotype as strains lacking the extra copy of the tRNA gene.

The approach was further developed when the plasmid PRI952 expressing both *dnaY* and the rare *ileX* tRNA gene was used to increase the expression of a flu antigen in *E. coli*¹¹³. Interestingly, the results show a slightly diminished level of expression when both *dnaY* and *ileX* were supplied on a plasmid, compared to merely *ileX*-supply. The authors do not go into further analysis of this observation, but for the work presented in chapter 2, this is highly relevant: We observe a similar and much more pronounced effect where the supply of several tRNA genes on a plasmid strongly diminishes the expression of a gene containing several rare codons. When the tRNA genes are not co-expressed, gene-expression is recovered.

The plasmid PRI952 was further developed and used for expression of parasite genes, which are known to have a high AT-content¹¹⁵. With the addition of *glyT* to the plasmid, creating pRIG, the supply of minor tRNAs cognate to rare arginine-, isoleucine-, and glycine-codons, improved the expression levels of the parasite-genes significantly¹¹⁵. The *glyT* gene was amplified from an *E. coli* genome: Although the source is not specified, but for both K-12 and BL21 (and derivatives), the region surrounding this gene is conserved. Our assesment of the sequences of the oligonucleotides used to amplify the *glyT* gene revealed that the other 3 tRNAs co-transcribed with *glyT* are also transferred to pRIG, and these tRNAs are very likely to play a role in the obtained results as well.

One of the first commercial tRNA-supplying plasmids, pRIL¹¹⁶, contains the tRNA genes *argU*, *ileY*, and *leuW*, and is supplied as an expression system in BL21(DE3). Curiously, unlike the K-12 MG1655 strain, BL21(DE3) has no *ileY* gene encoded in the genome, but the corresponding variant *ileX* is 97% identical, and carries the same anticodon. This plasmid has been used to study the effects of changing the concentration of tRNA on protein folding and solubility¹¹⁶. The results showed that the tRNAs are up-regulated between 5- and 27-fold, and additionally, that the concentrations of other genome-encoded tRNAs is not affected. The cell-wide study also shows how several proteins suffer from altered solubility, increasing the probability for them to end up aggregating, thus non-functional¹¹⁶.

The commercialization of the concept tRNA overexpression has been extended and several companies offer tRNA-supplying solutions, examples are in the CodonPlus and Rosetta strain collections. The strains in these collections carry a

plasmid supplying tRNAs for the most rare codons in *E. coli*, offering "universal expression" of eukaryotic genes^{68,117}.

In a recent large-scale protein production project, the Human Protein Atlas Project, changing the bacterial expression strain from the traditionally applied BL21 to the tRNA-supplying Rosetta improved the overall yields and success rate¹¹⁸. Following up on this work, the effect of different promoters in the two expression strains was tested, suggesting that Rosetta generally supported higher expression, regardless of promoter choice¹¹⁹. However, surprisingly, no correlation between (rare) codon content and effect of tRNA supply was observed in these studies¹¹⁹.

Paper I

This thesis provides a deeper understanding of the effects of the different side effects that may be encountered when attempting to increase protein production levels in bacteria. Chapter 2 reports the surprisingly small effects of supplying additional copies of low-abundance tRNA on a large set of membrane protein encoding genes. We expressed a library comprised of more than 300 *E. coli* inner membrane protein-encoding genes fused to GFP, to systematically study the effect of tRNA complementation in a high-throughput manner. To our surprise, the majority of the constructs showed little difference in expression with and without pRARE, and upon this discovery we initiated the search for genes that would be influenced by tRNA complementation. The results of expressing of a set of synonymous codon variants encoding a GFP-derived fluorescent protein caught us by surprise: Contrary to expectations, the presence of pRARE yielded a dramatic drop in expression levels. Upon construction of pRARE-derived plasmids either lacking all tRNAs or the T7 lysozyme gene, we show that the presence of additional tRNA copies can have a negative effect on heterologous gene expression, as no inhibitory effect on expression or growth was observed in absence of tRNAs on the pRARE backbone. Moreover in the absence of the *lysS* gene, the inhibitory affect was preserved in pRARE compared to pLysS. Still urging to investigate the positive effects of tRNA complementation, we noticed that several plant cytochrome P450 genes and a human breast cancer-related gene seemed to be positively affected in expression by the pRARE plasmid. Our initial experimental setup confirmed the positive effect of pRARE, but removal

of the tRNA encoding regions of pRARE resulted in a further increase in expression from both constructs. The *lysS* gene present on both pRARE and pLysS encodes T7 lysozyme. To the contribution of *lysS* to expression levels, expression of *lysS*-less plasmids revealed even higher expression levels than with the tRNA-less pRARE. This suggests that the major difference between the effect of pRARE and pLysS is caused by *lysS*. These phenomena seem to have been largely overlooked despite the huge popularity of the T7/pET-based systems for bacterial protein production.

1.12 Standardized cloning strategies

One of the key drivers for biological research has been the ability to cut and paste DNA in a controlled manner from different sources, and successively assemble these parts into gene constructs, also known as cloning. Advances in molecular biology are moving fast, but cloning remains a bottleneck, as it is time consuming and unpredictable. For production of complex compounds, the expressing of a single gene is often not sufficient, as the cells machinery do not provide all precursors. This is particularly a challenge when choosing unicellular organisms for cell factories such as *E. coli* and *S. cerevisiae*, as they often have a less versatile metabolism. To overcome the limitations of these hosts, introducing a full pathway for synthesis of the desired compound can be the solution. To introduce a whole pathway into a host strain, a great amount of cloning and subsequent optimization is required. Substantial efforts have been invested into developing new tools for DNA assembly, to increase the standardized format, and ensure improved, low-cost DNA assembly¹²⁰. With decreasing cost of gene synthesis, the cloning-bottleneck is minimized, and the advanced techniques for DNA assembly makes building libraries and larger multi-gene constructs even more feasible¹²¹.

With strategies like Gibson Assembly¹²¹ and Golden Gate cloning¹²², which both offers multi-fragment cloning in a single tube, it has become possible to tackle projects of increasing scale and complexity. Challenges still remain, as standardization often comes with the cost of flexibility, and to move into the future of

cell factories, the bottlenecks involved in cloning must be overcome, offering an even more generic standardized solution. We have addressed the paradox of sacrificing flexibility in exchange for optimized standardization in PaperII, which is presented here.

PaperII

In chapter 3, the work of Morten Thrane Nielsen, Susanna Seppälä, myself and others is presented, and it covers the development of a standardized method for multi-gene assembly. Production of the red compound protoporphyrin IX is applied as a proof-of-concept, due to the easy initial screening with a visible output. We have created a cloning strategy that allows for standardization at two levels. By defining an entry format and subsequent general downstream assembly strategy, the platform allows for faster cloning cycles. We utilized the cloning pipeline to explore aspects of pathway engineering in *E. coli* using protoporphyrin IX as test case. In two simple design-build-test cycles we managed to achieve state of the art levels of protoporphyrin IX production (50 mg/L). Besides easy assembly of genes, we optimized the level of expression through randomization of the Shine-Delgarno sequence, incorporating differentiated affinity in the ribosome binding site. The pipeline supports easy combination of standardized genes, and makes it easy to obtain full functioning pathways in *E. coli*, enabling a shorter optimization process.

1.13 High-value compound production in *E. coli*

To obtain full-length functional proteins and enzymes from heterologous expressed genes is another major challenge in the biotechnology. This can especially be challenging when they are membrane bound in their original host, as membrane properties varies greatly. Recent advances in recombinant protein production has allowed for the production of several high value products in *E. coli*, including several therapeutic proteins^{12,123}. However, the production of numerous valuable compounds is restricted by the successful expression of genes encoding important enzymes involved in their biosynthesis. These products include terpenoids, alkaloids and phenylpropanoids, which are often produced as secondary metabolites in plants. They

have applications in several fields such as human health and nutrition¹²⁴, and their complex structures are difficult to obtain through chemical synthesis. They often contain complex core carbon skeletons which are stereo specifically hydroxylized,, and this task is completed by cytochrome P450 monooxygenases (P450s)¹²⁴. The production of isoprenoids and likewise compounds in microbial hosts as an alternative to engineered plants is attractive, as they are much cheaper to grow, less susceptible to contamination, and downstream processing from microbial sources is usually easier¹²⁵. However, many obstacles must be overcome to achieve functional microbial expression of the genes responsible for biosynthesis of an isoprenoid of interest. The major limitation lies within the expression of P450s: so far, no reliable guidelines are developed for microbial P450 expression, and one of the challenges in obtaining functional enzymes are their putative dependency on membrane integration, as well as their dependency on redox partners¹²⁶.

The efficient expression of P450s requires a systematic assessment of the proper expression conditions. In Paper III we have contributed with a platform, from which several plant P450s have been expressed successfully in our hands.

Paper III

Chapter 4 contains the joined efforts of Ulla Christensen, Dario V. Albacete, myself and others, where we develop and test a platform for optimized expression of P450s in *E. coli*. Functional P450s are achieved by very simple means. The addition of a carboxy-terminal GFP yielded a visible out that could be easily assayed and optimized in several iterations, and the setup is very suitable for expression screening in a larger scale. We applied an N-terminal tag of 28 amino acids previously known to improve gene expression in *E. coli* and tested the effect of exchanging the P450s native membrane anchor for an anchor derived from an *E. coli* membrane protein gene. Additionally, we tested several growth media and strains, and determined a set of optimal conditions for P450 expression in *E. coli*. We show the efficiency of the platform by expressing six P450s originating from plants, and for all genes, we were able to achieve full-length, functional proteins. The work shows how a GFP-reporter approach can be useful for optimizing expression of P450s in *E. coli*, and shows that there is no inherent limitation for applying different types of strains.

1.14 Future perspectives on cell factories

Biotechnology has come a long way since the first cell factories were developed in the early 80's, and the fast-paced technology development decreases the time and money spent on cell factory development. One of the future perspectives for biotechnology is the technology to synthetically construct a fully functioning genome, which will allow for increased system control to increase production of a compound of interest. The initial steps towards synthetic genomes have already been taken: the genome of a small virus that infects bacteria was synthetically constructed in 2003¹²⁷, and the first synthetic yeast genome was revealed in 2014¹²⁸. This work emphasizes how easy and well-functioning standardized DNA assembly techniques have become, as undergraduate students constructed most of the DNA fragments. A new era of biology based on the redesign of genomes is emerging, and the way has been paved by this work. However, the heavy manual workload to construct synthetic genomes has to be dramatically reduced, and this will likely happen as more automated labor are introduced in laboratories. Additionally, when the price for DNA synthesis decreases dramatically, we will probably see a slowly emerging paradigm shift. It is unlikely that DNA cloning will be completely phased out; the flexibility, rapidness and accessibility is valued by scientists, and a shift in mindset will probably happen over several decades. Another future perspective is the expanded use of cell-free systems. The power of cell-free systems has been valued for more than 100 years¹²⁹, but has far from reached its full potential yet. By avoiding complex balancing of the cells and the scientist's objectives, the cellular resources can be focused towards an exclusive user-defined objective. Cell-free systems can be exploited to increase the understanding of Nature's way, and furthermore, the capabilities of cells can be exploited without having to nurture their needs. In the time between now, and until these and other advances are readily available for the majority of researchers, the classic way of developing cell factories is challenged all the time, i.e. through new cloning methods, strain design and better understanding of gene transfer between organisms. It is naive to believe that a complete shift to sustainable solutions can be achieved through biotechnology, but giant leaps in the right direction are taken, and this at least postpones the exhaustion of the planets resources.

1.15 References

1. Nations, U. Sustainable development. *Sustain. Dev.* (2015). at <http://www.un.org/sustainabledevelopment/development-agenda/>
2. Keasling, J. D. Manufacturing molecules through metabolic engineering. *Science* (80-.). **330**, 1355–1358 (2010).
3. Erickson, B., Nelson & Winters, P. Perspective on opportunities in industrial biotechnology in renewable chemicals. *Biotechnol. J.* **7**, 176–85 (2012).
4. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids. *Nature* **171**, 737–738 (1953).
5. Jackson, D. a, Symons, R. H. & Berg, P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 2904–2909 (1972).
6. Mertz, J. E. & Davis, R. W. Cleavage of DNA by R 1 restriction endonuclease generates cohesive ends. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 3370–3374 (1972).
7. Lobban, P. E. & Kaiser, A. D. Enzymatic end-to-end joining of DNA molecules. *J. Mol. Biol.* **78**, 453–471 (1973).
8. Cohen, S. N., Chang, a C., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3240–3244 (1973).
9. Ono, R. *Business of Biotechnology: From the Bench to the Street*. (Elsevier, 2013).
10. IBIS. Global Biotechnology Market Research Report | Sep 2015. *Glob. Biotechnol. Mark. Res. Rep. | Sep 2015* (2015). at <http://www.ibisworld.com/industry/global/global-biotechnology.html>
11. Palomares, L. A. & Kuri-breña, F. Industrial Recombinant Protein Production. *Biotechnology V*, - (2007).
12. Huang, C.-J., Lin, H. & Yang, X. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *J. Ind. Microbiol. Biotechnol.* **39**, 383–399 (2012).
13. Röhm, O. Patent DE 283923. (1913).
14. Nomoto, M. & Narahashi, Y. A Proteolytic Enzyme of *Streptomyces Griseus*: I. Purification of a Protease of *Streptomyces Griseus*. *J. Biochem.* **46**, 653–667 (1959).
15. Human insulin receives FDA approval. *FDA Drug Bull.* **12**, 18–9 (1982).
16. Lee, S. Y., Mattanovich, D. & Villaverde, A. Systems metabolic engineering, industrial biotechnology and microbial cell factories. *Microb. Cell Fact.* **11**,

- 156 (2012).
17. Demain, A. L. & Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* **27**, 297–306 (2009).
 18. Rosano, G. L. & Ceccarelli, E. a. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **5**, 1–17 (2014).
 19. Sahdev, S., Khattar, S. K. & Saini, K. S. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Mol. Cell. Biochem.* **307**, 249–64 (2008).
 20. Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E. & Villaverde, A. Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* **8**, 17 (2009).
 21. Westers, L., Westers, H. & Quax, W. J. *Bacillus subtilis* as cell factory for pharmaceutical proteins: A biotechnological approach to optimize the host organism. *Biochim. Biophys. Acta - Mol. Cell Res.* **1694**, 299–310 (2004).
 22. Morello, E. *et al.* *Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion. *J. Mol. Microbiol. Biotechnol.* **14**, 48–58 (2008).
 23. Punt, P. J. *et al.* Filamentous fungi as cell factories for heterologous protein production. *Trends Biotechnol.* **20**, 200–6 (2002).
 24. Petsch, D. & Anspach, F. B. Endotoxin removal from protein solutions. *J. Biotechnol.* **76**, 97–119 (2000).
 25. Palva, I. Molecular cloning of α -amylase gene from *Bacillus amyloliquefaciens* and its expression in *B. subtilis*. *Gene* **19**, 81–87 (1982).
 26. Bolotin, A. *et al.* The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res.* **11**, 731–53 (2001).
 27. Li, W. F., Zhou, X. X. & Lu, P. Bottlenecks in the expression and secretion of heterologous proteins in *Bacillus subtilis*. *Res. Microbiol.* **155**, 605–610 (2004).
 28. Brockmeier, U., Wendorff, M. & Eggert, T. Versatile expression and secretion vectors for *Bacillus subtilis*. *Curr. Microbiol.* **52**, 143–8 (2006).
 29. Oh, Y.-K., Palsson, B. O., Park, S. M., Schilling, C. H. & Mahadevan, R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* **282**, 28791–9 (2007).
 30. Schallmey, M., Singh, A. & Ward, O. P. Developments in the use of *Bacillus* species for industrial production. *Can. J. Microbiol.* **50**, 1–17 (2004).
 31. Mierau, I. *et al.* Industrial-scale production and purification of a heterologous protein in *Lactococcus lactis* using the nisin-controlled gene expression system NICE: the case of lysostaphin. *Microb. Cell Fact.* **4**, 15 (2005).
 32. Nikel, P. I. A brief guide to *Pseudomonas putida* as a microbial cell factory. *Bioessays* (2012). at <<http://philpapers.org/rec/NIKABG>>

33. Timmis, K. N. *Pseudomonas putida*: a cosmopolitan opportunist par excellence. *Environ. Microbiol.* **4**, 779–81 (2002).
34. *Encyclopedia of Life Sciences*. (John Wiley & Sons, Ltd, 2001).
35. Sezonov, G., Joseleau-Petit, D. & D'Ari, R. *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* **189**, 8746–9 (2007).
36. Lee, S. Y. High cell-density culture of *Escherichia coli*. *Trends Biotechnol.* **14**, 98–105 (1996).
37. Pope, B. & Kent, H. M. High efficiency 5 min transformation of *Escherichia coli*. *Nucleic Acids Res.* **24**, 536–7 (1996).
38. Kane, J. F. & Hartley, D. L. Formation of recombinant protein inclusion bodies in *Escherichia coli*. *Trends Biotechnol.* **6**, 95–101 (1988).
39. *Physiological Stress Responses in Bioprocesses*. **89**, (Springer Berlin Heidelberg, 2004).
40. Singh, S. M. & Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **99**, 303–10 (2005).
41. Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S. & Kim, J. F. Tracing Ancestors and Relatives of *Escherichia coli* B, and the Derivation of B Strains REL606 and BL21(DE3). *J. Mol. Biol.* **394**, 634–643 (2009).
42. LEDERBERG, J. & TATUM, E. L. Gene Recombination in *Escherichia Coli*. *Nature* **158**, 558–558 (1946).
43. H, M. J. EXPERIMENTS IN MOLECULAR GENETICS. *Cold Spring Harb. Lab.* (1972).
44. Smale, S. T. Beta-galactosidase assay. *Cold Spring Harb. Protoc.* **2010**, pdb.prot5423 (2010).
45. Studier, F. W. & Moffatt, B. a. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).
46. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).
47. Phillips, T. a., VanBogelen, R. a. & Neidhardt, F. C. Lon gene product of *Escherichia coli* is a heat-shock protein. *J. Bacteriol.* **159**, 283–287 (1984).
48. Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **72**, 211–22 (2006).
49. Gronenborn, B. Overproduction of phage lambda repressor under control of the lac promotor of *Escherichia coli*. *Mol. Gen. Genet.* **148**, 243–50 (1976).
50. Brosius, J., Erfle, M. & Storella, J. Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. *J. Biol. Chem.* **260**, 3539–41 (1985).
51. Elvin, C. M. *et al.* Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* **87**, 123–6 (1990).

52. Skerra, A. Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in *Escherichia coli*. *Gene* **151**, 131–5 (1994).
53. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–30 (1995).
54. Haldimann, A., Daniels, L. L. & Wanner, B. L. Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the *Escherichia coli* phosphate regulon. *J. Bacteriol.* **180**, 1277–86 (1998).
55. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–10 (1997).
56. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. at http://download.springer.com/static/pdf/942/art%253A10.1007%252Fs00253-006-0465-8.pdf?auth66=1418896599_c088d494f422e03775d47084187984be&ext=.pdf
57. Arditti, R. R., Scaife, J. G. & Beckwith, J. R. The nature of mutants in the lac promoter region. *J. Mol. Biol.* **38**, 421–6 (1968).
58. Chamberlin, M., McGrath, J. & Waskell, L. New RNA polymerase from *Escherichia coli* infected with bacteriophage T7. *Nature* **228**, 227–231 (1970).
59. Chamberlin, M. & Ring, J. Characterization of T7-specific ribonucleic acid polymerase: I. General properties of the enzymatic reaction and the template specificity of the enzyme. *J. Biol. Chem.* **248**, 2235–2244 (1973).
60. Moffatt, B. a & Studier, F. W. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**, 221–227 (1987).
61. McAllister, W. T., Morris, C., Rosenberg, A. H. & Studier, F. W. Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. *J. Mol. Biol.* **153**, 527–44 (1981).
62. Cheng, X., Zhang, X., Pflugrath, J. W. & Studier, F. W. The structure of bacteriophage T7 lysozyme, a zinc amidase and an inhibitor of T7 RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 4034–4038 (1994).
63. Wagner, S. *et al.* Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14371–6 (2008).
64. Ikemura, T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.* **2**, 13–34 (1985).
65. Ikemura, T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J. Mol. Biol.* **151**, 389–409 (1981).
66. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000 [In Process Citation]. *Nucleic Acids Res* **28**, 292 (2000).

67. Chen, D. & Texada, D. E. Low-usage codons and rare codons of *Escherichia coli*. *GENE Ther. Mol. Biol.* **10**, 1–12 (2006).
68. Novy, R., Drott, D., Yaeger, K. & Mierendorf, R. Overcoming the codon bias of *E. coli* for enhanced protein expression. *Innovations* 4–6 (2001).
69. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**, 255–8 (2009).
70. Marin, M. Folding at the rhythm of the rare codon beat. *Biotechnol. J.* **3**, 1047–1057 (2008).
71. Tsai, C. J. *et al.* Synonymous Mutations and Ribosome Stalling Can Lead to Altered Folding Pathways and Distinct Minima. *J. Mol. Biol.* **383**, 281–291 (2008).
72. Fredrick, K. & Ibba, M. How the sequence of a gene can tune its translation. *Cell* **141**, 227–229 (2010).
73. Zull, J. E. & Smith, S. K. Is genetic code redundancy related to retention of structural information in both DNA strands? *Trends Biochem. Sci.* **15**, 257–261 (1990).
74. Kurland, C. G. & Ehrenberg, M. Growth-optimizing accuracy of gene expression. *Annu. Rev. Biophys. Biophys. Chem.* **16**, 291–317 (1987).
75. Sharp, P. M. & Li, W. H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
76. Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. *Trends Biotechnol.* **22**, 346–353 (2004).
77. Wu, G. *et al.* SGDB: A database of synthetic genes re-designed for optimizing protein over-expression. *Nucleic Acids Res.* **35**, 76–79 (2007).
78. Welch, M. *et al.* Design parameters to control synthetic gene expression in *Escherichia coli*. *PLoS One* **4**, e7002 (2009).
79. dos Reis, M. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res.* **31**, 6976–6985 (2003).
80. Wright, F. The ‘effective number of codons’ used in a gene. *Gene* **87**, 23–29 (1990).
81. dos Reis, M., Savva, R. & Wernisch, L. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.* **32**, 5036–44 (2004).
82. Crick, F. H. Codon--anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**, 548–555 (1966).
83. Nelson, D. L. & Cox, M. M. *Lehninger Principles of Biochemistry 6th ed. Book* (2013).
84. Rudorf, S. & Lipowsky, R. Protein Synthesis in *E. coli*: Dependence of Codon-

- Specific Elongation on tRNA Concentration and Codon Usage. *PLoS One* **10**, e0134994 (2015).
85. Tuller, T. *et al.* An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–54 (2010).
 86. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **342**, 475–479 (2013).
 87. Mirzadeh, K. *et al.* Enhanced Protein Production in Escherichia coli by Optimization of Cloning Scars at the Vector-Coding Sequence Junction. *ACS Synth. Biol.* **4**, 959–65 (2015).
 88. Yokoyama, S. & Nishimura, S. in *Modif. Nucleosides Codon Recognit.* 207–223 (1995).
 89. Kirsebom, L. a & Svärd, S. G. Base pairing between Escherichia coli RNase P RNA and its substrate. *EMBO J.* **13**, 4870–6 (1994).
 90. Hartmann, R. K., Gößringer, M., Späth, B., Fischer, S. & Marchfelder, A. Chapter 8 The Making of tRNAs and More - RNase P and tRNase Z. *Prog. Mol. Biol. Transl. Sci.* **85**, 319–368 (2009).
 91. Blattner, F. R. The Complete Genome Sequence of Escherichia coli K-12. *Science* (80-.). **277**, 1453–1462 (1997).
 92. Zahler, N. H., Christian, E. L. & Harris, M. E. Recognition of the 5' leader of pre-tRNA substrates by the active site of ribonuclease P. *RNA* **9**, 734–745 (2003).
 93. Inokuchi, H. & Yamao, F. *tRNA*. (American Society of Microbiology, 1995). doi:10.1128/9781555818333
 94. Frank, D. N. & Pace, N. R. Ribonuclease P: unity and diversity in a tRNA processing ribozyme. *Annu. Rev. Biochem.* **67**, 153–80 (1998).
 95. Li, Z. & Deutscher, M. P. Maturation pathways for E. coli tRNA precursors: A random multienzyme process in vivo. *Cell* **86**, 503–512 (1996).
 96. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **26**, 5017–35 (1998).
 97. ALLEN, E. H., GLASSMAN, E. & SCHWEET, R. S. Incorporation of amino acids into ribonucleic acid. I. The role of activating enzymes. *J. Biol. Chem.* **235**, 1061–7 (1960).
 98. Murao, K., Tanabe, T., Ishii, F., Namiki, M. & Nishimura, S. Primary sequence of arginine transfer RNA from Escherichia coli. *Biochem. Biophys. Res. Commun.* **47**, 1332–7 (1972).
 99. Faulhammer, H. G. & Joshi, R. L. Structural features in aminoacyl-tRNAs required for recognition by elongation factor Tu. *FEBS Lett.* **217**, 203–11 (1987).
 100. Schuette, J.-C. *et al.* GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J.* **28**, 755–765 (2009).
 101. Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander D Johnson, Julian

- Lewis, Martin Raff, Keith Roberts, P. W. *Essential Cell Biology 4th ed.* (2013).
102. Keseler, I. M. *et al.* EcoCyc: Fusing model organism databases with systems biology. *Nucleic Acids Res.* **41**, 605–612 (2013).
 103. Grantham, R., Gautier, C., Gouy, M., Mercier, R. & Pavé, A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* **8**, 197 (1980).
 104. Bulmer, M. Coevolution of codon usage and transfer RNA abundance. *Nature* **325**, 728–730 (1987).
 105. Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA Abundance and Codon Usage in *Escherichia coli* at Different Growth Rates. *J. Mol. Biol.* **260**, 649–663 (1996).
 106. Dittmar, K. a, Sørensen, M. a, Elf, J., Ehrenberg, M. & Pan, T. Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Rep.* **6**, 151–157 (2005).
 107. Corporation, M. M. RosettaDE3 pLysS competent cells. at <[https://www.merckmillipore.com/DK/en/product/Rosetta™\(DE3\)pLysS-Competent-Cells,EMD_BIO-70956](https://www.merckmillipore.com/DK/en/product/Rosetta™(DE3)pLysS-Competent-Cells,EMD_BIO-70956)>
 108. Selzer, G., Som, T., Itoh, T. & Tomizawa, J. The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. *Cell* **32**, 119–129 (1983).
 109. Shaw, W. Chloramphenicol acetyltransferase: enzymology and molecular biology. *CRC Crit. Rev. Biochem.* **14**, 1–46 (1983).
 110. Ogura, Y. *et al.* Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic *Escherichia coli*. *Proc. Natl. Acad. Sci.* **106**, 17939–17944 (2009).
 111. Curran, J. F. Decoding with the A:I wobble pair is inefficient. *Nucleic Acids Res.* **23**, 683–688 (1995).
 112. Nakanishi, K. *et al.* Structural basis for lysidine formation by ATP pyrophosphatase accompanied by a lysine-specific loop and a tRNA-recognition domain. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7487–92 (2005).
 113. Del Tito, B. J. *et al.* Effects of a minor isoleucyl tRNA on heterologous protein translation in *Escherichia coli*. *J. Bacteriol.* **177**, 7086–91 (1995).
 114. Brinkmann, U., Mattes, R. E. & Buckel, P. High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the *dnaY* gene product. *Gene* **85**, 109–114 (1989).
 115. Baca, A. M. & Hol, W. G. J. Overcoming codon bias: A method for high-level overexpression of Plasmodium and other AT-rich parasite genes in *Escherichia coli*. *Int. J. Parasitol.* **30**, 113–118 (2000).
 116. Fedyunin, I. *et al.* tRNA concentration fine tunes protein solubility. *FEBS Lett.* **586**, 3336–3340 (2012).
 117. Carstens, C. BL21-CodonPlus™ Cells Correct Expression Problems Caused by Codon Bias. *STRATAGENE* **14**, (2001).

118. Tegel, H., Tourle, S., Ottosson, J. & Persson, A. Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta(DE3). *Protein Expr. Purif.* **69**, 159–67 (2010).
119. Tegel, H., Ottosson, J. & Hober, S. Enhancing the protein production levels in Escherichia coli with a strong promoter. *FEBS J.* **278**, 729–39 (2011).
120. Kosuri, S. & Church, G. M. Large-scale de novo DNA synthesis: technologies and applications. *Nat. Methods* **11**, 499–507 (2014).
121. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–5 (2009).
122. Engler, C. & Marillonnet, S. Golden Gate cloning. *Methods Mol. Biol.* **1116**, 119–31 (2014).
123. Kamionka, M. Engineering of therapeutic proteins production in Escherichia coli. *Curr. Pharm. Biotechnol.* **12**, 268–274 (2011).
124. Morant, M., Bak, S., Møller, B. L. & Werck-Reichhart, D. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Curr. Opin. Biotechnol.* **14**, 151–62 (2003).
125. Ferrer-Miralles, N. & Villaverde, A. Bacterial cell factories for recombinant protein production; expanding the catalogue. *Microb. Cell Fact.* **12**, 113 (2013).
126. Eugster, H. P., Bärtsch, S., Würigler, F. E. & Sengstag, C. Functional co-expression of human oxidoreductase and cytochrome P450 1A1 in *Saccharomyces cerevisiae* results in increased EROD activity. *Biochem Biophys Res Commun* **185**, 641–7 (1992).
127. Smith, H. O., Hutchison, C. A., Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: X174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci.* **100**, 15440–15445 (2003).
128. Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–8 (2014).
129. Buchner, E. & Rapp, R. Alkoholische Gährung ohne Hefezellen. *Berichte der Dtsch. Chem. Gesellschaft* **30**, 2668–2678 (1897).

Chapter 2

Different side effects of extra tRNA supply in typical bacterial protein production scenarios

Recombinant protein production is at the core of biotechnology and numerous molecular tools and bacterial strains have been developed over the past four decades for this purpose. A major challenge is the degeneracy of the genetic code, causing many genes to express poorly. One attempted generic solution is to supply with extra copies of low-abundance tRNAs to compensate for the presence of complementary rare codons in genes-of-interest. Here we show that such extra tRNA, supplied by the commonly used pLysSRARE2 plasmid, has very little effect on the production of a large number of membrane proteins in *E. coli*. Further, we observe two types of side effects: in some cases growth and gene expression is directly impaired by the extra tRNA sequences and in other cases the apparent positive effects are instead caused by a differential expression of the lysozyme gene encoded on the same plasmid. Next generation RNA sequencing of a strain containing the pLysSRARE2 plasmid reveal that several of the tRNA molecules are transcribed in the antisense orientation providing a possible explanation to the observed negative effects and the lack of rare codon complementation. These phenomena seem to have been largely overlooked despite the huge popularity of the T7/pET-based systems for bacterial protein production.

2.1 Introduction

The genetic code is degenerate and between one and six different codons can encode a single amino acid. The frequency for using a specific codon varies between different organisms - a phenomenon referred to as codon bias. This has long been recognized as a major challenge in heterologous gene expression, as the host might not transcribe and/or translate the foreign DNA sequence in an optimal manner. Partly, this may be caused by a parallel low abundance of tRNAs cognate to codons that are low in frequency – these codons are commonly referred to as rare codons¹.

Numerous studies have shown that synonymous codon mutations can have drastic consequences for gene expression. Factors such as mRNA stability and structure can be affected, but also initiation and elongation of translation, as well as protein folding²⁻⁵. It is possible that the redundancy in the genetic code have evolved as an extra layer of information stored in DNA – e.g. as a method to preserve structural information of proteins within the nucleotide content⁶. It is important to note that rare codons can be an advantage; the hypothesis that rare codons in the 5' end of genes may positively influence translation efficiency⁷ was recently corroborated by assaying the expression of 14.000 synthetic reporter genes with codon usage variation in the 5' end⁸.

Several strategies have been developed to optimize genes for expression (for recent reviews on the topic see ⁹⁻¹⁴), but the success rate is hard to gauge due to the inherent bias against negative results in scientific literature¹⁵. An attractive alternative to DNA re-coding is overexpression of low-abundance tRNA species. This way, the host, rather than the individual genes, is harmonized towards an ideal gene expression chassis. In one example, Brinckmann and co-workers coexpressed *dnaY* (now known as *argU*) with eukaryotic genes known to express poorly in *E. coli*¹⁶. The genes had a high content of AGA and AGG codons, which are recognized by tRNA_{AGG/AGA} produced from *dnaY*. Upon supplying *dnaY* on a plasmid, improved accumulation of product up to 30% of total cellular protein was achieved¹⁶. A variant of the approach was later used for expression of parasite genes, which are known to have a high AT-content, and addition of a plasmid supplying minor tRNAs cognate to rare arginine-, isoleucine-, and glycine-codons, improved the expression levels significantly¹⁷. The concept of tRNA overexpression has been further developed and commercialized, and several bacterial strains are available offering "universal expression" of eukaryotic

Chapter 2

genes¹⁸. In a recent large-scale protein production project, the Human Protein Atlas Project, changing the bacterial expression strain from the traditionally applied BL21 to the tRNA-supplying Rosetta improved the overall yields and success rate¹⁹. Following up on this work, the effect of different promoters in the two expression strains was tested, suggesting that Rosetta generally supported higher expression, regardless of promoter choice²⁰. However, curiously, no correlation between (rare) codon content and effect of tRNA supply was observed in these studies²⁰.

As with codon optimization, bias in the way science is published, makes it hard to evaluate the global success rate of an approach such as (rare) tRNA complementation. Here, we compare the homologous expression of a large set of membrane protein (MP) encoding genes and observe surprisingly small effects of supplying additional copies of low-abundance tRNA. Moreover, in searching for genes that are affected by tRNA supply, we find examples where a tRNA supplying plasmid have a direct negative effect on gene expression, and examples where other factors in the tRNA supplying plasmids are the main causative elements.

Abbreviations:

AU: Arbitrary units; MP: Membrane proteins; GFP: Green fluorescent protein

Results

Additional tRNA copies have no dramatic effect on homologous production of MPs in E. coli

The green fluorescent protein (GFP) is a highly useful reporter of proper folding and targeting of MPs¹⁵. When GFP is fused to the C-terminus of a MP, misfolding or mislocalization of the MP will effectively quench the fluorescence from the attached GFP. In contrast, a properly expressed MP-GFP, with the C-terminus in the cytoplasm (C_{in}), can be monitored by simple fluorescence detection techniques²¹. A library of nearly all *E. coli* multispanning MPs, C-terminally fused to GFP in a common pET-based expression vector system was previously developed by Daley and coworkers²². To systematically study the effect of tRNA complementation, we reduced this library to the extent that it finally contained only the C_{in} inner membrane proteins which made it compatible with a fluorescent readout of expression. We then transformed the individual clones into both BL21(DE3) pLysS and the BL21-derived strain RosettaTM2(DE3) that harbours the pLysSRARE2, a rare tRNA-supplementing plasmid derived from pLysS²³ (Fig. 1a), and assayed protein production by whole cell fluorescence in a 96-deep well format.

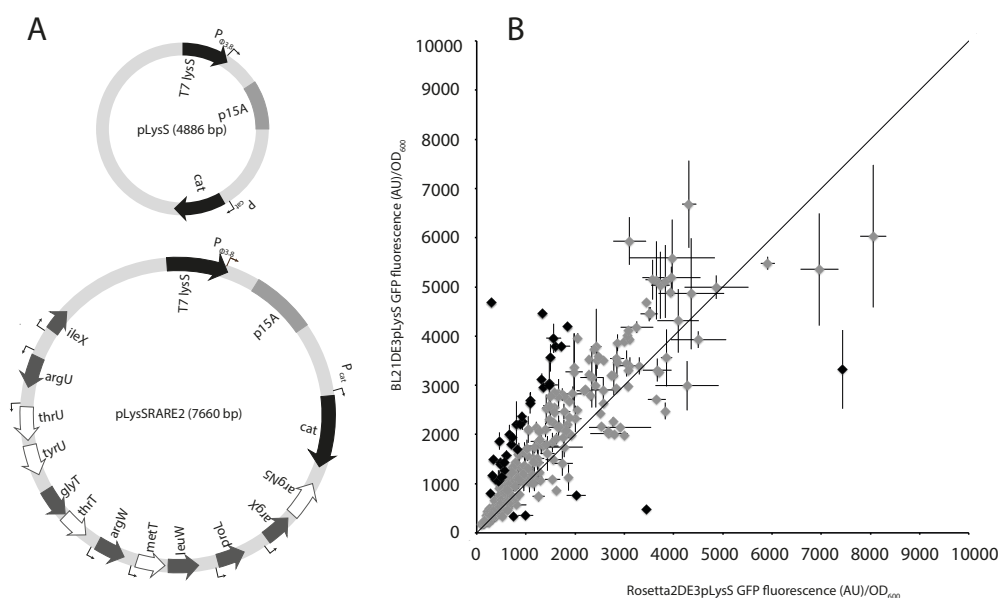


Figure 1. A) Schematic representation of pLysS and pRARE, not drawn to scale. On both plasmids: T7 *lysS*, $P_{\Phi 3.8}$, *cat*. On pRARE: 7 rare tRNAs (grey) and 5 non-rare tRNAs (white). B) Compared expression of ~300 membrane protein encoding genes in Rosetta2(DE3)pLysS on the X-axis and BL21(DE3)pLysS on the Y-axis. Units are fluorescence (arbitrary units)/OD_{600nm}.

To our surprise, the majority of the constructs showed little difference in protein production in the two strains (Fig. 1b): approximately 10% of the MPs were produced more than 2-fold in RosettaTM2(DE3). In contrast, only less than 2 % of the MP output was more than 2-fold better in BL21(DE3) pLysS. Apparently, tRNA complementation with pLysSRARE2 has no dramatic effect on overproduction of homologous MPs in *E. coli*.

Additional tRNA copies can have a negative impact on heterologous gene expression

In searching for genes that would be influenced by tRNA complementation, we came across a set of synonymous codon variants, encoding a synthetic GFP variant named DasherGFP (excitation and emission wave length at 510nm and 521nm respectively), expressed from a still inducible but reduced version of the rhamnose promoter prhaBAD. These variants were designed by scientists at DNA2.0, Inc. (Claes Gustafsson and Mark Welch, personal communication) to explore the effect of rare codons on protein production. In contrast to our expectations, several of the tested DasherGFP variants showed a remarkable drop in their production in presence of pLysSRARE2 in a number of different expression strains and growth conditions. We decided to study this further and focused on the DasherGFP25 variant that contains a total of 13 rare codons (out of 258 codons in total). In a representative experiment, we co-transformed the DasherGFP25 construct together with pLysSRARE2 or the pLysS control into BL21(DE3). By diluting an overnight culture, we started all expression cultures from OD₆₀₀ 0.05 and let them grow at 37°C to their exponential phase (OD 0.3 – 0.8) where the expression of DasherGFP25 was induced by adding 2 mM L-rhamnose. After 4 hours of induction, protein production levels were estimated by whole cell GFP fluorescence in a plate reader. As observed in our pilot experiments, the presence of pLysSRARE2 had a major inhibitory effect on produced DasherGFP25, showing approximately 15-fold less production under these conditions (Fig. 2a). Additionally, we observed that the growth rate of the expression strain was significantly affected under these conditions, suggesting a fitness cost caused by the presence of the pLysSRARE2 plasmid.

Despite the overall similarity with pLysS (Fig. 1a), apart from the tRNA genes, pLysSRARE2 contains additional elements included in the tRNA-encoding region of the plasmid that could influence heterologous gene expression or the bacterial cell physiology. In order to proof this and for having a control for the effect of tRNAs, we deleted all tRNA coding sequences from the plasmid (pLysSRARE2-tRNAless), creating a pLysS-alike plasmid, and co-transformed this construct with the DasherGFP25 construct. In the absence of all tRNAs, no inhibitory effect on protein production or growth was observed (Fig. 2a). To test the effect of the *lysS* gene, we also clone a pLysSRARE2 version without the *lysS* gene (pLysSRARE2-*lysS*less). Here, in the absence of *lysS*, the inhibitory affect was preserved compared to pLysSRARE2-tRNAless and pLysS (Fig. 2a). We conclude, the presence of additional tRNA copies can have a negative effect on heterologous protein production.

Elements other than tRNAs in pLysSRARE2 may play a significant role in boosting heterologous gene expression

We broadened our search for genes relying on pLysSRARE2 for proper expression and noticed that several plant cytochrome P450 genes and a human breast cancer-related gene fragment BRCA1 (Alex Toftgaard Nielsen and Ariane Zutz, personal communication) seemed to be positively affected in expression by the pLysSRARE2 plasmid. Here, we focus on the P450 CYP720B4 from the plant *Pinus sitchensis* involved in the biosynthesis of resin acid, a secondary metabolite involved in the defense against insects²⁴ and a fragment of BRCA1, a protein that plays a role in maintaining genomic stability and acts as a tumor suppressor.

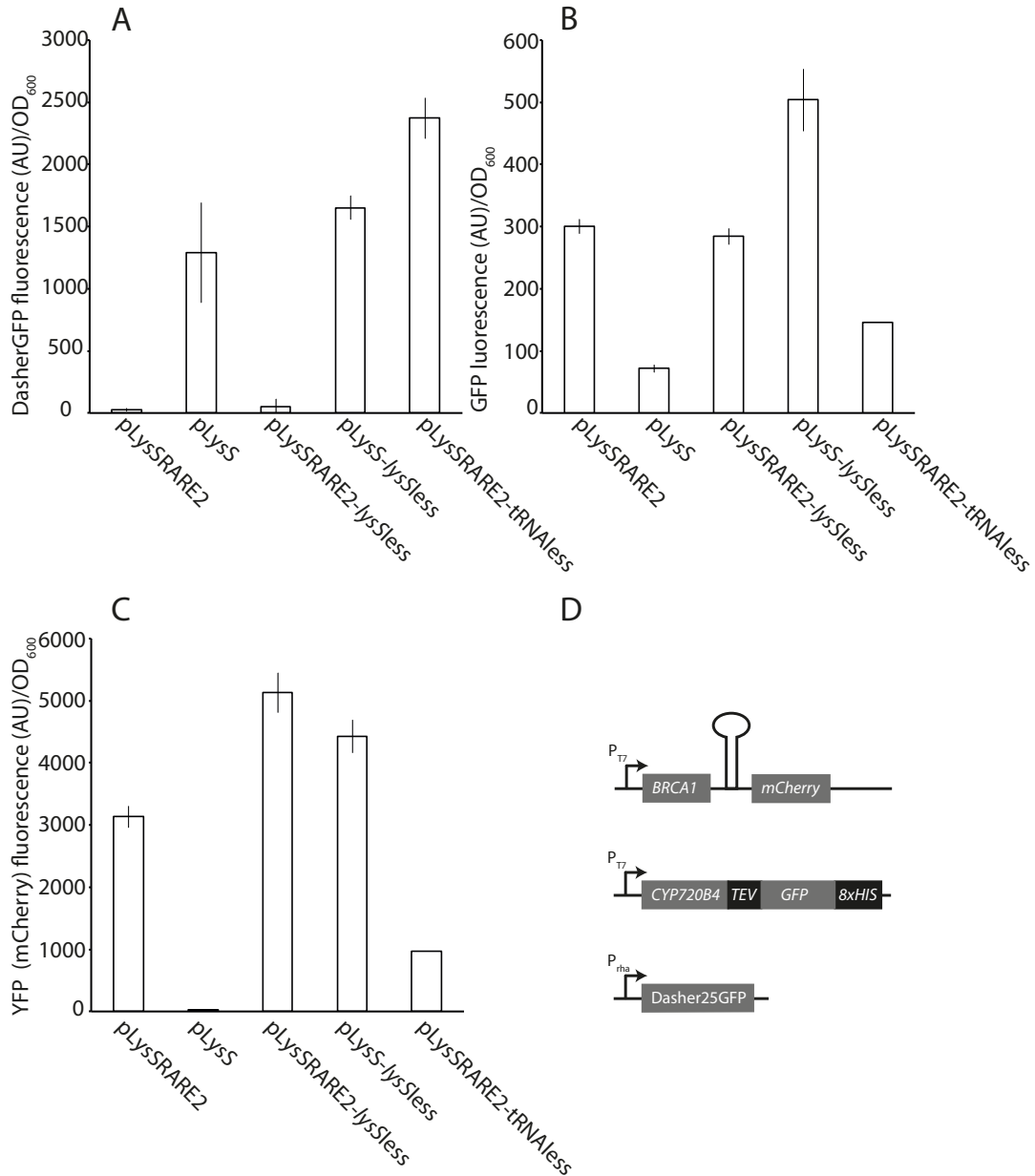


Figure 2. A) Fluorescence (arbitrary units)/OD_{600nm} barcharts for A) dasher25, b) CYP720B4 and C) BRCA1 expressed in pRARE, pLysS, pRARE-LysS, pLysS-LysS and pRARE-tRNAs. D) Schematic representation of constructs of P_{T7} -*BRCA1* translationally coupled to *mCherry*, P_{T7} -*CYP720B4* linked to *gfp*, and P_{rha} -*dasher25*

Both expression constructs are based on the common T7 promoter, figure 2d, and the CYP720B4 membrane protein is fused to GFP in constructs identical to those in the described *E. coli* MP library, whereas proper *BRCA1* translation can be followed using fluorescence due to a translationally coupled red fluorescent protein. Our initial experimental setup confirmed the positive effect of pLysSRARE2; CYP720B4 and BRCA1 expressed to 4- and 130-fold higher levels, respectively, compared to the

pLysS-controls (Figure 2b and 2c). However, to our surprise, removal of the tRNA encoding regions of pLysSRARE2 (the pLysSRARE2-notRNAs plasmid) resulted in a further increase in expression from both constructs. Thus, it appears that the positive effect of pLysSRARE2 could not be attributed to the extra tRNA copies. T7 lysozyme, encoded by the *lysS* gene on both pLysSRARE2 and pLysS, inhibits T7 polymerase²⁵, and thus could be a contributing factor in the differential expression of genes controlled by the T7 promoter. As an additional control we therefore assayed *CYP720B4* and *BRCA1* expression with pLysSRARE2 and pLysS variants where we had deleted the *lysS* gene. With the *lysS*-less pLysS variant, *CYP720B4* expressed at even higher levels than with pLysSRARE2-tRNAless suggesting that the major difference between the effect of pLysSRARE2 and pLysS was caused by *lysS* (Figure 2b). With *BRCA1*, the effect of *LysS* was also apparent as the expression increased more than 150-fold upon removing *lysS* from pLysS. Despite the fact that pLysSRARE2 is directly derived from pLysS, preserving the entire *lysS* gene and surrounding sequence, including the Φ 3.8 promoter that drives expression of *lysS*²³ it is not surprising that *lysS*-effects differ between the two plasmids. The Φ 3.8 promoter is placed *downstream* from the *lysS* gene and thus T7 RNA polymerase will only transcribe *lysS* after having transcribed the entire plasmid sequence present upstream from *lysS*²³. In the case of pLysSRARE2, this includes all the tRNAs sequences (Figure 1a) and a total of 2774 nucleotides in addition to those present in the corresponding region in pLysS. To confirm the difference in expression of *lysS* from pLysS and pLysSRARE2, we probed the presence of T7 lysozyme by semi-quantitative western blotting and observed that it was nearly completely absent from cells harbouring the pLysSRARE2 plasmid compared to pLysS (Figure 3a). To probe the functionality and consequences of the pLysSRARE2 plasmid more broadly, we expressed a *dasher* gene for 4 hours in BL21DE3 with pLysSRARE2, or pLysS, and isolated and sequenced the total RNA from these two strains. Although many genes appeared to express differently in the two strains, we were unable to identify clear trends of differentially expressed genes characteristic for specific physiological responses such as stress or starvation (Supplementary Information Table S1). Since T7 RNA polymerase needs to transcribe the majority of the pLysSRARE2 plasmid to reach the *lysS* gene, it could have a major impact on the way the tRNAs are transcribed from the plasmid too. Interestingly, the RNA sequencing analysis showed

Chapter 2

that the two tRNAs placed immediately downstream from the $\Phi 3.8$ promoter and the *cam* resistance gene (*argX* and *argN5*) were expressed to very high levels compared to the pLysS strain (Figure 3b). This seems to confirm that tRNA genes are over expressed from the pLysSRARE2 plasmid. However, when analysing the strand specificity of the sequencing, more than 99% of the RNA sequence originated from the antisense strand of *argX* and *argN5* and, thus, will not lead to functional tRNA molecules.

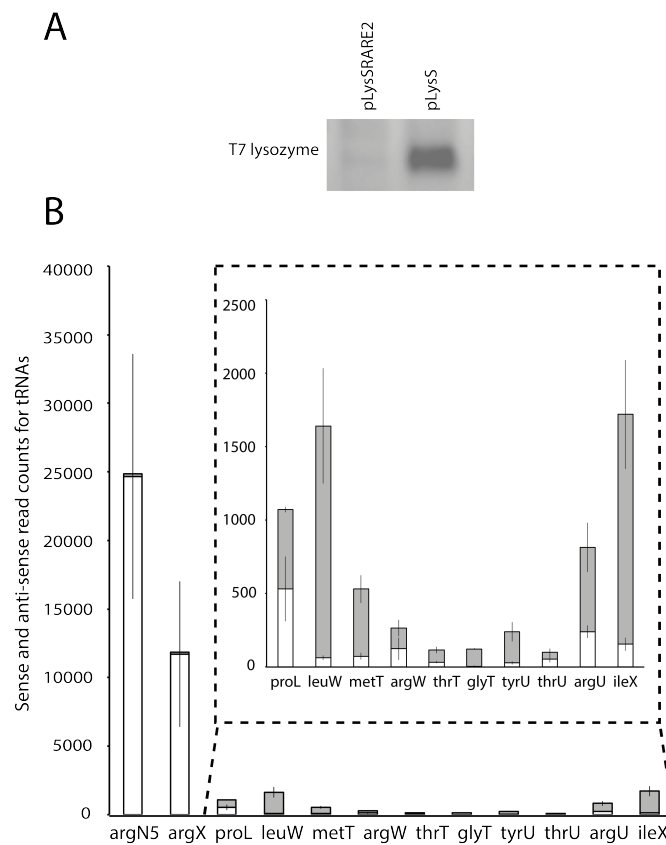


Figure 3. A) Immunoblot of T7 lysozyme for pLysSRARE2 and pLysS. B) Sense (grey) and anti-sense (white) read counts from RNA sequencing for all tRNA genes from pLysSRARE2.

The tRNA genes downstream from *argX* and *argN5* were not expressed to the same high levels and the strand specificity was not as prominent, although the significant proportion of sequence from the antisense strand of several of the other tRNAs is worth noting (Figure 3b).

2.2 Discussion

It has previously been noted that tRNAs are post-transcriptionally modified and amino-acetylated, and these processes must be essential for complementation of low-abundance tRNAs and could have metabolic side effects in the cell²⁶. Indeed, Fedyunin and co-workers showed that when expression of specific tRNAs was upregulated between 5- and 27-fold, several proteins in the cell suffered from altered solubility, making them aggregation-prone²⁷. It is likely that the apparent negative of impact of pLysSRARE2 on growth and on expression of *dasher* is related to such metabolic side effects, although our RNAseq analysis gave no specific clues in this direction. Instead, our analysis seems to point to an apparent design flaw in the pLysSRARE plasmid: T7 RNA polymerase transcribes major parts of the plasmid from the Φ 3.8 promoter, leading to antisense versions of several of the encoded tRNAs. *ArgX* recognizes CGA and CGG – two of the most rarely occurring codons in *E. coli*²⁸ – and it appears that heavy over expression of the antisense version of this tRNA can only make things worse. As showed here, another consequence of the pLysSRARE2-LysS design is that *lysS* is expressed at much lower levels than in pLysS. This is probably a direct consequence of the extra DNA added between the Φ 3.8 promoter and *lysS* in pLysSRARE2. In line with this, we observe in the RNAseq data that the levels of (antisense) tRNA decrease with the distance to the Φ 3.8 promoter (Figure 3b) Given that T7 lysozyme directly inhibits T7 RNA polymerase, it is therefore no surprise that genes expressed from the T7 promoter express very differently when combined with either pLysS or pLysSRARE2 – as shown here for two very different genes and expression constructs. Given the huge popularity of the T7/pET-based expression systems, it is obviously important to know if the developed tools work as intend. There still seems to be very little direct proof available for the efficiency of tRNA overexpression in providing “universal expression”.

2.3 Materials and methods

Bacterial strains

Escherichia coli strain NEB 5-alpha (New England BioLabs, Ipswich, USA) was used for cloning of PCR products and propagation of plasmids. The following *E. coli* strains were used for gene expression: RosettaTM2(DE3) pLysS and BL21(DE3)

pLysS (Novagen, Merck Millipore, Germany), homemade BL21(DE3) competent cells

DNA constructs

tRNA genes were removed from pRare using the following USER oligonucleotide sets

5'AGTCGGTGUCATCCTGGAACCTCGGCTACCTGATTTTC3'

3'ACACCGACUAAGATGTGCTGATATAGCTCAGTTGGTAGAG5'

5'ATCGATUCCAGCGGCCGCAAAGGCT3'

3'AATCGAUTGGATTATAAAGTAACTCCGTGTC5'

5'ATCATCGAUCAATGCGGCATGAGTATACCCGCTAATG3'

3'ACACCGACUTTACAATTCAATCAGTTACGCCTTCTTTATATCCTC5'

5'AGTCGGTGUCGTTGATACCGCAATGCGGTGTAATC3'

3'ATCGATGAUATTTTGAACCCCGCTTCGGCGGGGTTTTTTG5'

and *T7 lysS* was removed from pRARE and pLysS using the following USER oligonucleotides

5'ATATACGUGGTTTCCCTTTGTTCGCATTGGA3'

3'ACGTATAUTATTTCTTTCCTCCTTTCCTTTTTTAATCTATCAAA5'

followed by assembly with uracil excision technology as previously described (Nour-Eldin et al., 2006; Nørholm, 2010)^{28,29}.

PCR and DNA assembly

PCR products were amplified in 50 µL reactions containing: 1 µL homemade PfuX7 DNA polymerase, 0.2 mM dNTPs (Thermo Scientific, Waltham, USA), 1 mM MgCl₂, 1 mM DMSO, 0.5 µM forward oligonucleotide, 0.5 µM reverse oligonucleotide, Phusion® HF Reaction Buffer (New England BioLabs, Ipswich, USA) and 50 ng plasmid template. A touch-down PCR program was used for amplification: Step 1: 2 min 98°C; step 2: 15 sec 98°C, 20 sec 65°C (-1°C per cycle), 45 sec per kb at 72°C (step 2 repeated 9 times until 55°C, then repeated 20 cycles at the annealing temperature 55°C); step 3: 5 min 72°C; step 4 hold at 10°C. PCR products were gel purified using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany) and eluted in 10% TE buffer. A Nanodrop

Chapter 2

spectrophotometer 2000 (Thermo Scientific, Waltham, USA) was used for estimation of PCR product and vector concentration. Purified PCR products were incubated with 1 μ L USERTM enzyme (New England BioLabs, Ipswich, USA) for 30 min at 37°C and subsequently mixed with linearized vector backbone. 3 μ L of the assembled PCR product:vector solution was transformed into NEB 5-alpha chemically competent cells according to the manufacturer's protocol. Transformants were selected on Luria Bertoni (LB) agar plates supplemented with 25 μ g/mL kanamycin, 17 μ g/mL chloramphenicol. Colonies were screened for gene insert by colony PCR using OneTaq 2X Master mix (New England BioLabs, Ipswich, USA). Vectors were extracted and purified using QIAprep Spin Kit (Qiagen) and verified by sequencing.

Culture medium and expression conditions

All experiments were performed with biological triplicates with BL21(DE3) as expression host. Overnight (ON) cultures in 96-deep well plates: 500 μ L (DasherGFP variants BRCA1) or 800 μ L (CYP720B4) terrific broth (TB) medium supplemented with 25 μ g/mL kanamycin and 17 μ g/mL chloramphenicol was inoculated with single colonies and grown overnight at 37°C (DasherGFP variants, BRCA1) or 30°C (CYP720B4), at 300 rpm in a 5cm orbital shaking incubator. ON cultures were diluted to a start OD_{600nm} of 0.05 in fresh TB medium including antibiotics. The expression culture volume in 96-deep well plates was 500 μ L (Dashers, BRCA1), in 24-deep well plates it was 5 mL (CYP720B4); all grown to an OD_{600nm} of 0.3-0.5. After induction with 2 mM L- rhamnose (DasherGFP variants) or 0.4 mM IPTG (BRCA1, CYP720B4), growth was continued for 4 hours at 37°C, 300 rpm (DasherGFP variants) or overnight at 25°C, 300 rpm (BRCA1) or 3 hours 25°C, 150 rpm (CYP720B4).

Whole cell fluorescence measurements

Whole cell fluorescence was measured using 200 μ L (Dashers and BRCA1) or 2 mL (CYP720B4) induced culture. CYP720B4 cultures were harvested (2,500xg, 20 min) and the pellets resuspended in a total volume of 100 μ L PBS buffer. The GFP fluorophore was allowed to form for 2 h at room temperature (RT). The cultures harbouring the DasherGFP variants or BRCA1 were measured directly after culturing. Fluorescence was detected by using excitation/emission wave length depending on the construct: 505nm/525nm for the DasherGFP variants,

Chapter 2

580nm/610nm for BRCA1-mCherry and 485nm/512nm for CYP720B4-GFP with a window of +/- 9 nm, gain value 50 (CYP720B4) or 80 (BRCA1), using plate reader SynergyMx SMATLD (BioTek, Winooski, USA).

SDS-page

Cell pellets from 450 uL expression culture were resuspended in 20 mM Tris-HCL pH 8.0 including benzonase (0.5uL/mL) and adjusted to equal OD values. Cell lysis was done by freeze-thawing and the cell lysate was subsequently mixed 1:1 with SDS-sample buffer (125 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 100 mM DTT, bromphenol blue) and boiled at 99 °C for 5 minutes. For SDS-PAGE, equal volumes of cell lysate-sample buffer mixes were loaded and run in parallel with PageRuler™ Prestained Protein Ladder 10-170K (Thermo Scientific, Waltham, USA) on RunBlue 4-20% gradient SDS gels (Expedeon, USA) with RunBlue Rapid running buffer (Expedeon, USA) at 125 kV for 60 mins.

Immunoblotting

Proteins were transferred from SDS-PAGE onto nitrocellulose sheets (iBlot transfer stacks with Novex mini nitrocellulose membranes and iBlot Dry blotting system,,ThermoFischer Scientific, USA). Blocking was done ON at 4 °C with 5% (w/v) dried milk in TBS-T buffer, followed by 4x5mins washing with excessive TBS-T volumes. Incubation with the 1° antibody (anti-T7 lysozyme from rabbit, 1:5000 dilution) for 1 h at RT in TBS-T solution was followed by excessive washing with TBS-T, 1h incubation with 2° antibody (anti-rabbit from donkey, 1:2500 dilution) in TBS-T and again excessive washing. Detection was performed with Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare Life Sciences, UK).

RNA isolation

1 mL samples in triplicates were harvested after 4h induction and snapfrozen at -80C after STOP solution (5% Phenol for RNA, 100% EtOH) treatment. Total RNA isolation was performed using Qiagen RNAeasy kit according to the manufacturers' instructions. The protocol includes DNase treatment.

RNA library preparation and sequencing

The sequencing libraries were prepared in triplicates using a TruSeq RNA Stranded Sample Preparation kit (Illumina Inc., San Diego CA) and pooled together prior to

sequencing. An average cDNA library size was determined using the Agilent DNA 1000 kit on an Agilent 2100 Bioanalyzer. Libraries were normalized and pooled in 10 mM Tris-Cl, pH 8.0, plus 0.05% Tween 20 to the final concentration of 10 nM. Denaturated in 0.2N NaOH, 10 pm pool of 6 libraries in 600 ul ice-cold HT1 buffer was loaded onto the flow cell provided in the MiSeq Reagent kit v2 300 cycles and sequenced on a MiSeq® (Illumina Inc., San Diego CA) platform with a paired-end protocol and read lengths of 151 nt. On a regular basis, concentration and quality of the nucleic acids were determined by using a Qubit 2.0 fluorometer (Invitrogen) or an Agilent 2100 Bioanalyzer (Agilent Technologies).

Analysis of RNA-sequencing data

TopHat (2.1.0) and Cufflinks (2.2.1) suite was used for RNA-seq-based differential gene expression analysis as described by Trapnell, C. *et al*³⁰. Reference genome and annotations for BL21(DE3) strain were retrieved from NCBI Reference Sequence Database (Version NC_012971.2). Features present on the pDR861-SR, pLysS and pRARE plasmids are added to reference manually. Sense and anti-sense read counts for tRNAs present on pRARE are determined using HTSeq library described by Anders, S. *et al*³¹ over reads mapped to plasmids alone.

Acknowledgements

This work was supported by The Novo Nordisk Foundation.

2.4 References

1. Jakubowski, H. & Goldman, E. Quantities of individual aminoacyl-tRNA families and their turnover in Escherichia coli. *J. Bacteriol.* **158**, 769–776 (1984).
2. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in Escherichia coli. *Science* **324**, 255–8 (2009).
3. Marin, M. Folding at the rhythm of the rare codon beat. *Biotechnol. J.* **3**, 1047–1057 (2008).
4. Tsai, C. J. *et al.* Synonymous Mutations and Ribosome Stalling Can Lead to

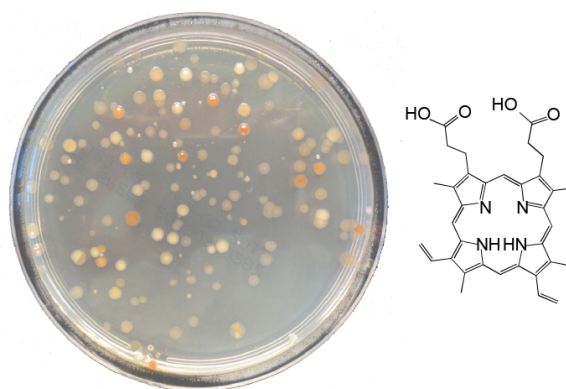
- Altered Folding Pathways and Distinct Minima. *J. Mol. Biol.* **383**, 281–291 (2008).
5. Fredrick, K. & Ibba, M. How the sequence of a gene can tune its translation. *Cell* **141**, 227–229 (2010).
 6. Zull, J. E. & Smith, S. K. Is genetic code redundancy related to retention of structural information in both DNA strands? *Trends Biochem. Sci.* **15**, 257–261 (1990).
 7. Tuller, T. *et al.* An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* **141**, 344–54 (2010).
 8. Goodman, D. B., Church, G. M. & Kosuri, S. Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science* **342**, 475–479 (2013).
 9. Rosano, G. L. & Ceccarelli, E. a. Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Front. Microbiol.* **5**, 1–17 (2014).
 10. Makrides, S. C. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60**, 512–538 (1996).
 11. Baneyx, F. Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10**, 411–421 (1999).
 12. van de Linde, S. *et al.* Investigating cellular structures at the nanoscale with organic fluorophores. *Chem. Biol.* **20**, 8–18 (2013).
 13. Jana, S. & Deb, J. K. Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **67**, 289–298 (2005).
 14. Sørensen, H. P. & Mortensen, K. K. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **115**, 113–128 (2005).
 15. Nørholm, M. H. H. *et al.* Manipulating the genetic code for membrane protein production: What have we learnt so far? *Biochim. Biophys. Acta - Biomembr.* **1818**, 1091–1096 (2012).
 16. Brinkmann, U., Mattes, R. E. & Buckel, P. High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the

- dnaY* gene product. *Gene* **85**, 109–114 (1989).
17. Overcoming codon bias: A method for high-level overexpression of Plasmodium and other AT-rich parasite genes in Escherichia coli. at <http://dtu-ftp.cvt.dk/cgi-bin/fulltext/elsevier?pi=/0020/7519/v0030i02/00001340.pdf&key=83447271>
 18. Novy, R., Drott, D., Yaeger, K. & Mierendorf, R. Overcoming the codon bias of E. coli for enhanced protein expression. *Innovations* 4–6 (2001).
 19. Tegel, H., Tourle, S., Ottosson, J. & Persson, A. Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta(DE3). *Protein Expr. Purif.* **69**, 159–67 (2010).
 20. Tegel, H., Ottosson, J. & Hober, S. Enhancing the protein production levels in Escherichia coli with a strong promoter. *FEBS J.* **278**, 729–39 (2011).
 21. Daley, D. O. *et al.* Global topology analysis of the Escherichia coli inner membrane proteome. *Science* **308**, 1321–3 (2005).
 22. Daley, D. O. *et al.* Global topology analysis of the Escherichia coli inner membrane proteome. *Science* **308**, 1321–3 (2005).
 23. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60–89 (1990).
 24. Hamberger, B., Ohnishi, T., Hamberger, B., Séguin, A. & Bohlmann, J. Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiol.* **157**, 1677–95 (2011).
 25. Moffatt, B. a & Studier, F. W. T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell* **49**, 221–227 (1987).
 26. Welch, M. *et al.* Design parameters to control synthetic gene expression in Escherichia coli. *PLoS One* **4**, e7002 (2009).
 27. Fedyunin, I. *et al.* tRNA concentration fine tunes protein solubility. *FEBS Lett.* **586**, 3336–3340 (2012).

28. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000 [In Process Citation]. *Nucleic Acids Res* **28**, 292 (2000).
29. Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K. & Halkier, B. A. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* **34**, e122 (2006).
30. Nørholm, M. H. H. A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol.* **10**, 21 (2010).
31. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–78 (2012).
32. Anders, S., Pyl, P. T. & Huber, W. HTSeq A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2014).

Chapter 3

Assembly of highly standardized gene fragments for high-level production of porphyrins in *E. coli*



Standardization of molecular cloning greatly facilitates advanced DNA engineering, parts sharing and collaborative efforts like the iGEM competition. All of these attributes facilitate exploitation of the wealth of genetic information made available by genome and RNA sequencing. Standardization also comes at the cost of reduced flexibility. We addressed this paradox by formulating a set of design principles aimed at maximizing standardization while maintaining high flexibility in choice of cloning technique and minimizing the impact of standard sequences. The design principles were applied to formulate a molecular cloning pipeline and iteratively assemble and optimize a six-gene pathway for protoporphyrin IX synthesis in *Escherichia coli*. State of the art production levels were achieved through two simple cycles of engineering and screening. The principles defined here are generally applicable and simplifies the experimental design of projects aimed at biosynthetic pathway construction or engineering.

3.1 Introduction

The ability to recombine DNA with restriction enzymes and ligase moved molecular biology into a new era and was pivotal in the birth of biotechnology. More recently, whole genome sequencing has revolutionized molecular biology and paved the way for fields like metabolic engineering and synthetic biology. Exploitation of the unprecedented availability of sequence information seems now mainly limited by our ability to build, test and analyze variants of constructs, genes and pathways.

Accordingly, a range of advanced methods for DNA cloning have recently emerged, addressing the more complex demands ¹⁻⁵. Moreover, the need to standardize molecular cloning and share biological parts have been strongly emphasized ⁶. A prime example is the biobrick standard proposed by Tom Knight ⁷ that was followed by a parts sharing initiative heavily promoted by the iGEM synthetic biology competition. Several alternative biobrick formats and technologies have subsequently been described (BBF RFC process.

http://openwetware.org/wiki/The_BioBricks_Foundation: RFC). However, the history of biobricks exposes two inherent complications in the pursuit of standardizing molecular cloning; 1) newly developed molecular cloning methodologies are not *per se* backward compatible, which compromises the benefits of part sharing and 2) standardization invariably results in addition of short sequences mediating the standardized DNA assembly. We decided to face these challenges, by formulating a set of design principles to guide the construction of a standardized DNA assembly pipeline. With the proposed design principles, the level of standardization is maximized while maintaining high flexibility in choice of cloning technique and minimizing the impact of standard sequences. The approach incorporates a substantial degree of variability and opportunities for custom design. As test case for our cloning pipeline, we selected production of protoporphyrin IX in *Escherichia coli*.

Protoporphyrins have several properties that make them attractive as a model system; 1) the red color of protoporphyrins allow for an easy optical assay, 2) the fluorescent properties of protoporphyrins allow specific and sensitive detection and relative quantitative comparison by HPLC, 3) Authentic standards are available for several protoporphyrins and 4) the enzymes involved in protoporphyrin biosynthesis are thoroughly characterized ⁸. Protoporphyrin IX is ubiquitously distributed in nature as a precursor in the synthesis of essential prosthetic tetraprotoporphyrin groups such as

in b- and c-type cytochromes, peroxidases, chlorophylls, cytochrome P450s and haemoglobins. Moreover, protoporphyrins, particularly the manganese containing ones, have application potential within the biomedical field as drug precursors and modulators of cancer cell growth ⁹.

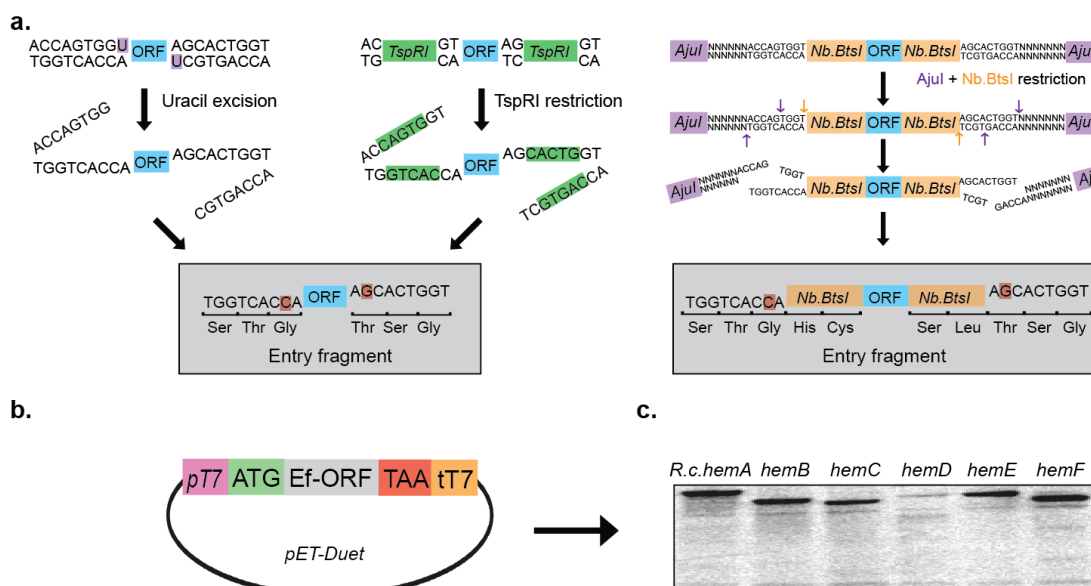


Figure 1. Activation of entry format fragments. The entry format is defined by the 9-bp single stranded overhang (sso) specified in the design principles. The 9-bp can be formed by a number of PCR and restriction enzyme based techniques. Panel a, Left: PCR based technique where uracils are incorporated in oligo nucleotides and sso's are generated by enzymatic removal of the uracils after the PCR. The uracils mediating sso formation are highlighted in purple. Middle panel: TspRI restriction. Right panel: Combined type-II-S restriction endonuclease (here AjuI) and nicking endonuclease (here Nb.BtsI). A consequence of this approach is that the recognition sites for the nicking endonuclease are also translated resulting in addition of two amino acids on both sides of the ORF. The resulting entry fragments with processed sso's are illustrated below. One base pair difference in the upstream and downstream sso's ensures processed directional cloning (highlighted in red). The amino acids encoded by the 9 bp entry format sso are displayed below the nucleotide sequence. Panel b; illustration of the entry vector with entry-ORF inserted. The entry vector adds a T7 promoter, a start codon, a stop codon and a T7 terminator to the cloned ORF (these elements are highlighted in pink, green, red and orange respectively). Panel c; results of T7 dependent ³⁵S-methionine labeling of heme biosynthetic genes *hemA-F*.

Protoporphyrins are synthesized from glycine and succinyl-CoA through the Shemin pathway ^{10,11} or from tRNA loaded with Glutamate through the C5-pathway ¹²⁻¹⁴. d-Amino levulinic acid (ALA) is the first common intermediate in the two pathways. *E. coli* naturally utilize the C5-pathway ¹⁵ and introduction of the Shemin pathway by heterologous expression of d-amino levulinic acid synthase (*hemA*) ^{16,17} has previously been exploited to obtain high ALA titers ¹⁸. In this manuscript, we describe a standardized yet generic pipeline for DNA cloning and demonstrate its feasibility for pathway engineering by constructing a pair of plasmids that enable high

production of protoporphyrin IX (80 mM). The plasmids are selected through two engineering cycles based on screening of pathway variants, balancing expression of the six biosynthetic genes with semi-randomized Shine-Dalgarno (SD) sequences.

3.2 Results and Discussion

Design principles for cloning pipeline:

The overall purpose of our cloning pipeline is to achieve streamlined, robust, high-throughput cloning with a minimal requirement for *de novo* design. In addition, the pipeline design, and underlying principles, should be sufficiently generic to be compatible with a range of advanced cloning techniques. The major principles that we arrive at to fulfill these criteria is to standardize at two levels; 1) define an entry format and corresponding entry vector to achieve fully standardized incorporation of new DNA-parts and 2) design a general strategy for downstream parts assembly. The overall design principles that we arrive at for the entry format is that all entry-parts will be flanked by defined 9-bp sequences that facilitate directional ligase independent cloning with both PCR and restriction enzyme based technologies (Figure 1a). Specifically, the 9-bp flanking sequences are designed to contain recognition sequences for TspRI. TspRI is a unique restriction enzyme that generates a 9-bp single stranded overhang, thus providing a one step, error free method for activation prior to cloning (Figure 1a). The major advantage of this approach is that synthetic genes, free from internal TspRI sites, can be sub-cloned in a PCR-error-free manner. The flexible entry format can also be created with simple PCR-based methods such as uracil-excision^{19,20}, enzyme free cloning²¹ and PliCing⁵. These methodologies will likely be useful when genes are cloned from the native source DNA. As a third alternative compatible entry fragments can be created using a methodology that resembles type-II-S restriction enzyme-based technologies such as MoClo²² and Golden Gate²³. This third alternative relies on addition of internal 6-bp *Nb. BtsI* nicking enzyme recognition sites and is therefore ligase independent. Type-II-S restriction enzyme-based technologies may be particular useful in robotics-based systems. Homology/recombination-based cloning methodologies such as isothermal in vitro recombination¹, ligase independent cloning²⁴ or ligase cycling reaction²⁵ could also be use the generate entry clones since these are highly flexible and sequence independent technologies. The second principle is that protein coding

sequences are designed without start and stop codons. This ensures flexibility in terms of modifications such as addition of tags, linkers or fused gene constructs. A consequence of this is that the 9-bp sequences that facilitate cloning will be translated and could potentially interfere with protein function. To minimize this risk, the TspRI-containing 9-bp sequences are designed to encode only Serine, Threonine and Glycine. Serine and Glycine are the most commonly used residues in protein linker sequences²⁶ and Threonine only differ from Serine by one methyl group.

Finally, we decided to utilize a T7-based expression vector as entry plasmid backbone. This introduces a simple quality checkpoint for proper translation of protein encoding entry fragments based on an assay for specific T7-dependent expression in the presence of rifampicin²⁷. In our case, we utilized a pET-DUET-1 backbone as entry vector. The backbone was modified to position a start codon in optimal distance from a consensus SD sequence and immediately upstream of the 9-bp entry sequence to ensure that no unwanted amino acids were included in the resulting protein. In principle any backbone can be prepared for hosting biobricks in this format by a plethora of techniques including PCR, TspRI or type-II-S restriction enzymes (Figure S1).

Application of the standardized cloning pipeline to protoporphyrin IX production

With the basic principles established we decided to utilize the cloning pipeline to explore aspects of pathway engineering in *E. coli* using protoporphyrin IX as test case. We introduced the Shemin pathway in *E. coli* by heterologous expression of the model ALA synthase, *hemA*, from *Rhodobacter capsulatus* as previously described^{18,28}. The remaining enzymes required for protoporphyrin IX biosynthesis, encoded by *hemB*, *hemC*, *hemD*, *hemE* and *hemF*⁸, were acquired from *E. coli*.

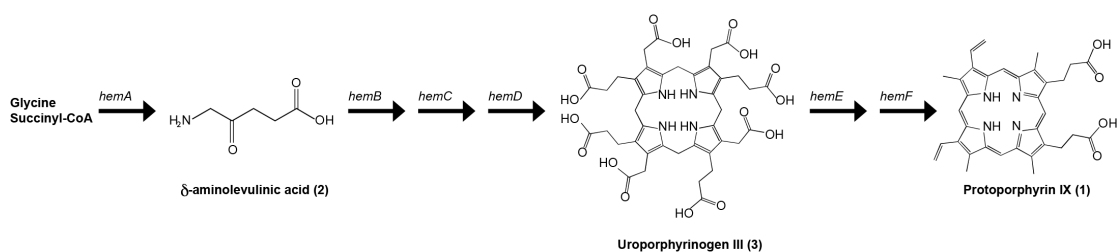


Figure 2. Biosynthetic route for protoporphyrin IX by the Shemin pathway. Chemical structures of the expected products of *hemD*, uroporphyrinogen III and *hemF*, protoporphyrin IX as well as the first dedicated intermediate d-aminolevulinic acid are displayed.

Chapter 3

A schematic overview of the biosynthetic pathway is presented in Figure 2. All six genes were inserted in the pET-Duet-1 based entry vector as entry fragments by uracil-excision and transformed in *E. coli* BL21 DE3 for verification of transcription and translation by ^{35}S -methionine labeling (Figure 1b).

Indeed, translation of all six genes was confirmed. *hemD* appeared to be translated at a lower level indicating a possible bottleneck in pathway engineering (Figure 1c).

Standardized assembly of multiple fragments.

The next step in our pipeline construction was to standardize assembly of several fragments. Minimal *de novo* design, to avoid DNA assembly being the experimental bottleneck, and compatibility with multiple techniques and flexibility in terms of number of assembled fragments were the drivers for our design principles.

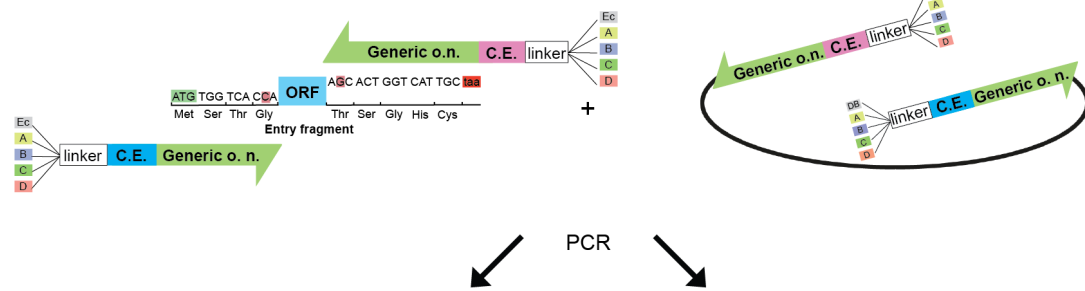
Table 1. Generic linkers used for multi-fragment assembly

Linker ID	Sequence (sense)	Complementary sequence (anti-sense)	Length (bp)
Entry fragment upstream	ACCAGTGGT	ACCACTGGT	9
Entry fragment downstream	AGCACTGGT	ACCAGTGCT	9
Generic linker A	AGTCGGTGT	ACACCGACT	9
Generic linker B	AAGCAGCGT	ACGCTGCTT	9
Generic linker C	AGACGTCAT	ATGACGTCT	9
Generic linker D	AGGTCTGAGT	ACTCAGACCT	10

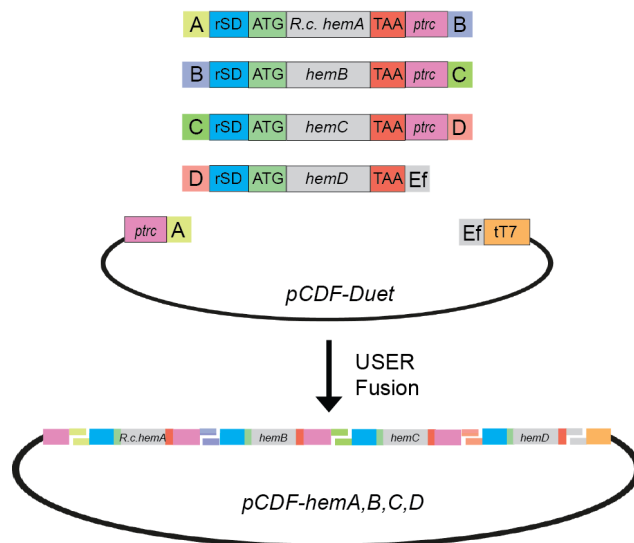
At this stage, we focused on PCR based techniques as they allow highly flexible assembly of multiple fragments in a single cloning reaction^{1,5,20}. We arrived at the following simple design rules; 1) entry fragments should be PCR amplified exclusively by generic oligo nucleotides, which is possible due to the standardized entry formats allowing generic oligo nucleotides annealing in the common Gly-, Ser-, and Thr-encoding regions, and 2) linking regions that define specificity in multi-

fragment assemblies should be designed with maximal sequence diversity and confined length (8-10bp) (Table 1). These principles allow generation of a very large number of assemblies using only a small set of benchmarked oligo nucleotides, thus minimizing *de novo* optimization and failed assemblies due to erroneous oligo nucleotides.

a.



b.



c.

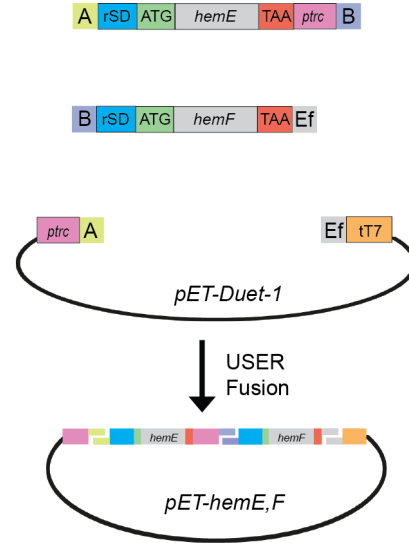


Figure 3. Schematic overview of multi-fragment assembly from entry clones. Panel a; Illustration of entry clone (left) and backbone (right) amplification by generic oligo nucleotides. The three elements of the generic oligo nucleotides are depicted as follows; annealing part (green), control elements (blue upstream, purple downstream) and linker (white box). The five variants of the linker (A-D, Entry fragment (Ef)) are shown in extension of the oligo nucleotide. The schematic overviews are not drawn to scale.

The approach ensures backward compatibility as long as the linkers are maintained, making it straightforward to expand the assembly at a later stage. Moreover, the short

overlap sequences allow for incorporation of control elements (e.g. required for transcription and translation) while keeping the oligo nucleotides relatively short. This means that the gene expression levels can be tailored using permutations of the generic oligo nucleotides with e.g. variable promoters, SD-sequences, leader sequences or peptide tags. In summary, generic oligo nucleotides contain three defined regions: An annealing part, a control element and an assembly linker (Figure 3a). In contrast to the entry clone strategy, the described second-level multigene assembly strategy comes at the expense of flexibility in terms of cloning technology.

Assembly of tetrapyrrole core biosynthesis genes

The proof of concept pathway engineering was conducted as a screening study with constitutive promoters (*ptrc*) and diversity at the translational level mediated by semi-randomized SD-sequences (NNNNNGGAN). The core GGA motif in the SD was fixed to reduce the occurrence of non-functional sequences as previously described²⁹. The semi-random SD was included as a control element in the upstream- and *ptrc* in the downstream-oligo nucleotide for the entry clone, while a *ptrc* and a transcriptional terminator was included as control elements in the backbone oligo nucleotides (Figure 3a). Entry clones of *R. capsulatus hemA*, *hemB*, *hemC* and *hemD* were PCR amplified individually with the linker pairs A-B, B-C, C-D and D-Ec and assembled with a PCR amplified *pCDF* backbone harboring the Ec-A linkers (Figure 3b) and transformed into *E. coli* DH5a.

Chapter 3

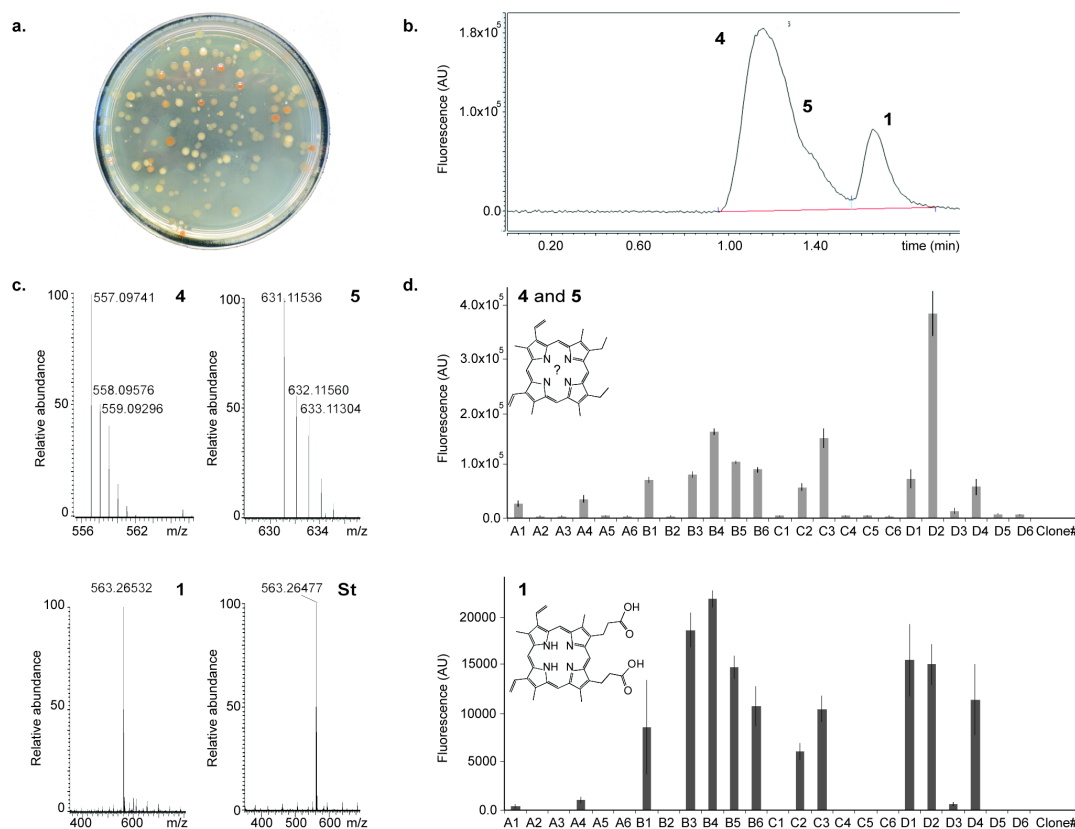


Figure 4. Protoporphyrin production in stage-one (hemA-D) clones. Panel a; colonies were screened by optical inspection directly on plates 72 h after transformation. Protoporphyrin producing clones were identified based on color intensity and 24 clones spanning the observed color intensity range were selected for analysis. Panel b; representative HPLC-FLD chromatogram of protoporphyrin producing clone (D1). The observed compounds are numbered 1,4 and 5. Panel c; representative mass spectra of protoporphyrin producing clone (D2). The panels show the dominant ions in each of the compounds 1,4 and 5 as well as an authentic standard (St) of protoporphyrin IX. Panel d; quantitative analysis of protoporphyrin production in 24 stage-one clones based on HPLC-FLD data. Each clone was analyzed in seven replicates and the graphs depict the average signal intensities with standard deviations. The upper panel summarizes the levels of 4 and 5 combined, as they could not be separated with the HPLC parameters used in this study. A general tetrapyrrole structure is embedded to illustrate that the exact structure of the compounds is unknown, but the fluorescence and mass spectrometrical data indicate a protoporphyrin class molecule. The lower panel depicts the protoporphyrin IX levels and the chemical structure of protoporphyrin IX (Pubchem) is embedded.

The initial screen for protoporphyrin producing transformants was conducted directly on the LB-plate used for transformation. Simple optical inspection revealed highly diverse levels of protoporphyrin production among the transformants (Figure 4a). Approximately 10% of the colonies were intensely red-colored, 10% were white and the remaining had an intermediate color phenotype. Twenty-four *E. coli* colonies spanning the range of protoporphyrin producers were selected for further analysis. The *pCDF-hemAD* plasmids were purified and sequenced to establish the efficiency of the uracil-excision cloning and map the variance in SD-sequences.

Chapter 3

The success rate of the cloning, defined as clones containing all four genes in correct order, was 79% (19 out of 24 clones as verified by sequencing).

Table 2. Shine Dalgarno sequences from analyzed *pCDF-hemA-D* clones, their predicted RBS strengths and protoporphyrin production levels

<i>pCDF-hemA-D plasmids</i>					Predicted Shine-Dalgarno strength				Protoporphyrin (sum of compound 4 and 5)
Clone ID	<i>hemA</i> SD	<i>hemB</i> SD	<i>hemC</i> SD	<i>hemD</i> SD	<i>hemA</i> SD	<i>hemB</i> SD	<i>hemC</i> SD	<i>hemD</i> SD	
A1	GCATTGGAG	TGTACGGAA	TCTTTGGAT	CTCCGGGAG	535	238	489	312	26906
A2	GGTATGGAT	GGTCCGGAT	CCGGAGGAT	TCATTGGAG	228	139	8514	586	3478
A3	AATTCGGAG	ATTTCGGAA	n.d.	CAATGGGAA	1507	586	125	511	3547
A4	TGCCCCGAG	GTGAGGGAT	GCGTGGGAT	CCGCCGGAT	641	229	2065	120	35769
A5	GACATGGAT	CGCGCGGAC	ATTTTGGAG	ACGTTGGAA	260	326	9452	51	4079
A6	AAGGGGGAT	AAGATGGAT	GTGGTGGAT	CCCTAGGAT	2526	103	1843	208	3447
B1	GTGATGGAG	GAGATGGAA	ACCGCGGAG	CGTTGGGAG	878	278	767	373	71167
B2	TGGTCGGAC	GGACTGGAG	AGTACGGAT	TTTTTGGAC	97	373	1440	373	3873
B3	AAGAGGGAT	-	-	-	518				81363
B4	ACCATGGAG	GACTTGGAT	ATGTTGGAT	CGTTAGGAG	767	85	802	1440	160586
B5	TCCTTGGAG	GTTCCGGAT	GTTCCGGAG	TGAGTGGAA	1725	174	4600	670	104191
B6	ATTGTGGAG	TGGCAGGAG	CGCTAGGAG	GTTGGGGAT	918	560	2705	97	90191
C1	GCTCAGGAT	TGTTGGGAA	CTGGAGGAT	AGCCTGGAA	52	447	5429	228	4196
C2	TTCGTGGAG	CGTTGGGAG	TCCTTGGAG	CCCTAGGAG	1259	670	2160	2586	57429
C3	CTCCAGGAT	ACGTTGGAG	TTGTAGGAT	TGTGGGGAT	139	139	2065	116	148578
C4	GGATTGGAT	TCGGAGGAG	GTGGTGGAG	GATGTGGAT	238	1440	4642	14	4028
C5	TTTTCGGAT	TCATTGGAG	__TCAGGAG	TTGAGGGAT	260	373	6304	470	4009
C6	CGAGCGGAT	_CGCGGGAG	ATCACGGAG	TCCTTGGAT	50	918	2065	133	3191
D1	GTTTTGGAG	-	-	-	1051				73002
D2	TATTGGGAG	AGACGGGAT	GTGGGGGAT	GAGTTGGAG	1772	285	13354	560	380934
D3	GTAGGGGAA	AGGGGGGAT	TTTTTGGAT	CCCGTGGAT	429	939	1576	139	13363
D4	GTAGAGGAT	-	-	-	802				58055
D5	-	-	-	TAAGTGGAC				312	7656
D6	-	-	-	GTCTCGGAC				68	6492

The five remaining clones contained only one gene, which was either *hemA* (3 clones) or *hemD* (2 clones). This was not surprising, as these genes contain one linker with perfect match to the vector. The predicted strength of the 24 different combinations of SD-sequences were calculated using the The Ribosome Binding Site Calculator^{30,31} (Table 2), but no statistically significant correlations between predicted RBS strength and protoporphyrin levels were observed.

Chemical characterization of stage-one clones

All 24 selected *E. coli* colonies were subjected to detailed analyses by chromatography-based methods. HPLC coupled fluorescence detection (FLD) using excitation and emission wavelength selective for protoporphyrins (410nm excitation/620nm emission) revealed the presence of several signals (Figure 4b). Identification of the fluorescent compounds was attempted with Fourier transform mass spectrometry (FT-MS) analysis. The late eluting compound, **1**, was identified as protoporphyrin IX based on both the retention time (1.6 min) and comparison with the mass spectra of an authentic standard (Figure 4c). The major signal obtained, however, turned out to contain two related co-eluting compounds **4** and **5** sharing an ion of 1169 m/z and with dominant ions of 577 and 631 m/z, respectively (Figure S2). Neither of these ions matches those derived from expected intermediates. The FT-MS analyses support that both **4** and **5** contain nitrogen and are capable of binding iron in the high molecular mass form (1169 m/z) and the FLD signal indicate a tetrapyrrole structure. Taken together with the exclusive co-occurrence of protoporphyrin IX with elevated levels of **4** and **5** (Figure 4d), this evidence suggests that **4** and **5** are indeed protoporphyrins. A metabolite with similar mass as **1** (557 m/z) has previously been observed in co-occurrence with protoporphyrin IX in *E. coli* constitutively expressing *R. capsulatus hemA*³². The metabolite was identified as heme based on proposed signature ions, however, the accurate mass and isotope patterns available to us demonstrate that iron is not bound in the low molecular form (557 m/z), but exclusively in the high molecular form (1167 m/z) which rules heme out as a target. Previous studies identified uroporphyrin I and III as the dominant species of protoporphyrins upon constitutive expression of *R. capsulatus hemA* with *E. coli hemB*, *C* and *D*²⁸. However, in our experiments, neither uroporphyrin I nor III was detectable in any of the analyzed clones (data not shown). Absence of

uroporphyrin III could be explained by insufficient *hemD* activity as the *in vitro* translation data suggested that *hemD* was not as efficiently expressed as the remaining genes. The presence of unknown tetrapyrrols are therefore best explained by conversion of protoporphyrin intermediates into unknown derivatives and protoporphyrin IX by endogenous *E. coli* enzymes.

Quantitative analysis of stage-one clones

Based on the data from the HPLC-FLD analyses, we performed a relative quantification of the protoporphyrin producing capabilities of the clones in our collection (Figure 4d). Approximately half of the clones produced trace amounts of **4** and **5** with no detectable protoporphyrin IX and are thus considered the background level. The remaining clones produced 4-53 fold more **4** and **5** than the background level while the protoporphyrin IX levels were fairly constant. The three clones containing only *hemA* (B3, D1 and D4) all produced protoporphyrin IX as well as decent amounts of **4** and **5** (Figure 4d), thus supporting previous observations that the initial step of heme biosynthesis indeed is limiting in *E. coli*¹⁸. From a pathway engineering perspective, the clone D2 was by far the most promising as its ability to produce **4** and **5** was superior thus possibly offering a substantial gain in product formation upon re-engineering with the remaining pathway genes (Figure 4d). To verify that the protoporphyrin producing phenotype of clone D2 is indeed facilitated by the *pCDF-hemA,B,C,D* plasmid, the purified plasmid was re-transformed into DH5a cells. Six clones were randomly chosen for analysis as described above. All six produced both protoporphyrins in similar levels as the original D2 clone thus confirming that the *pCDF-hemA,B,C,D-D2* plasmid is responsible for this phenotype (data not shown).

Building on stage-one clones for increased protoporphyrin IX production

In a second engineering cycle entry fragments of *hemE* and *hemF* were amplified with matching generic oligo nucleotides (A-B, B-Ec), assembled with a pET-DUET backbone (Ec-A) and transformed into DH5a harboring *pCDF-hemA,B,C,D-D2* (Figure 3c). Again, transformants were selected based on color intensity and two superior clones c15 and c16 emerged from the subsequent HPLC-FLD screen.

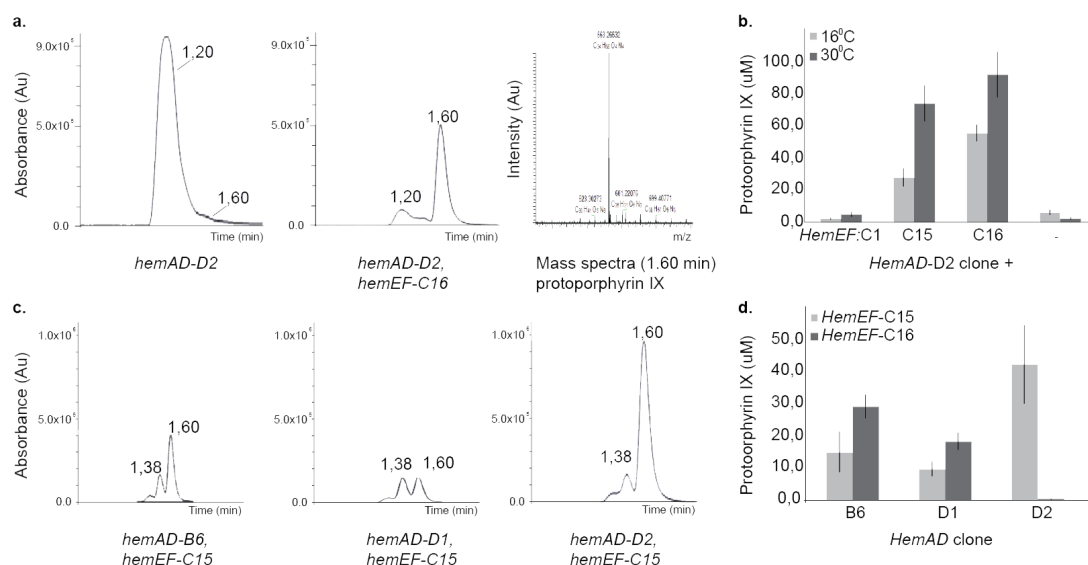


Figure 5. Protoporphyrin IX production in stage-two (*hemAD*+*hemEF*) clones. Panel a; HPLC-FLD chromatograms of the parental stage-one clone (*hemAD-D2*, left) and the highest producer among the stage-two clones, *hemAD-D2, hemEF-c16* (middle) as well as mass spectrum of protoporphyrin IX produced in *hemAD-D2, hemEF-c16*. (right). Panel b absolute quantitation of protoporphyrin IX produced by stage-two clones at 16 °C (light gray) or 30 °C (dark gray). Bars display the average of three replicates and error bars denote the standard deviation. Panel c HPLC-FLD chromatograms from three stage-one clones re-transformed with the high producing stage-two plasmid *pET-hemEF-c15*. Panel d absolute quantification of protoporphyrin IX produced by three stage-one clones re-transformed with either *pET-hemEF-c15* (light gray) or *pET-hemEF-c16* (dark gray) cultivated at 30 °C. Bars display the average of eight replicates and error bars denote the standard deviation.

Both clones displayed a dramatic increase in protoporphyrin IX production accompanied by an equally dramatic decrease in the accumulation of **4** and **5**, thus further supporting that **4** and **5** are intermediates in or derivatives of intermediates in protoporphyrin IX synthesis (Figure 5a). Absolute quantification of protoporphyrin IX by LC-FTMS revealed production capabilities ranging from 10-90uM corresponding to 6 to 50 mg/l (Figure 5b). Thus the best producing clones obtained (*pCDF-hemAD-D2*+*pET-hemEF-c15* or *C16*) are within the same range as the highest producers previously reported (84+/-8 mM)²⁸. To further investigate the impact of the *pET-hemE,F* plasmids from c15 and c16 on protoporphyrin production, the plasmids were isolated, purified, sequenced and re-transformed into three stage-one clones: an average producer (B6), a single gene *hemA* clone (D1) and the high producer (D2). In five of six combinations, transformation with *pET-hemE,F* resulted in increased levels of protoporphyrin IX and reduced accumulation of **4** and **5** (Figure 5c-d). Curiously, the last combination (*pCDF-hemAD-D2, pET-hemEF c16*) resulted in clones that did not produce detectable levels of protoporphyrins as they did in the initial screen and in the identical replicate. To investigate whether this phenotype was

caused by alterations at the DNA level, *pCDF-hemAD-D2* and *pET-hemEF-c16* were isolated individually from a transformant randomly chosen among the analyzed colonies, and compared to the corresponding plasmid stocks by restriction analysis. Indeed, the *pCDF-hemAD-D2* plasmid displayed an altered DNA fragment pattern, consistent with loss of *R. capsulatus hemA*, (Figure S3) and upon re-examination of the 24 stage-one clones we found several cases of instability over the prolonged incubation time of the assay. In the light of the induced plasmid instability observed in this study, and the previously obtained titers in the same range²⁸, it is tempting to speculate that production of protoporphyrins in the observed range is near the maximum that *E. coli* can tolerate.

3.3 Concluding remarks

In summary, we have demonstrated that by utilizing the cloning pipeline and design principles proposed in this manuscript, state of the art levels of protoporphyrin IX production (50 mg/l) can be achieved in *E. coli* through two simple design-build-test cycles. Moreover, the design allows not only maximizing production, but also generation of intermediary producers by alternative plasmid combinations. Although we do not obtain higher titers than previously shown²⁸ our work demonstrates that standardized genes and proteins (all six enzymes in the pathway contained the same linker sequences) can perform as well as the native versions. In addition to the example described here, we have recently successfully added the same peptide linkers to seven genes and proteins in a heterologous pathway for high-level production of a di-terpene in *E. coli* (Nielsen, MT and Nørholm, MHH, unpublished). The incorporated flexibility makes this approach useful for optimization of production scenarios as well as for rapid generation and iterative expansion of collections of clones with varying degrees of a defined phenotype. In the simplified design used for our proof of principle studies, we introduced a risk for genetic instability by repeated use of the *ptrc* promoter. This may explain the observed occurrence of instable plasmids. For construction of stable production strains, this risk can trivially be eliminated by introducing different promoters as control elements in each set of generic oligo nucleotides. As the assembly of fragments relies exclusively on generic oligo nucleotides, part amplification can be heavily benchmarked and requires little *de novo* optimization. When combined with a robust cloning technique, such as the uracil excision cloning used in this manuscript (79% success rate with five DNA

fragments), screening based approaches become feasible and are readily accommodated in automated processes. Taken together, the simple design process, flexibility for fine-tuning by tailoring control elements in generic primers, reliable amplification and robust cloning, combined with the potential for automation, makes the pipeline a very useful tool for combinatorial assembly and optimization of biosynthetic pathways.

3.4 Methods

Strains and media

Escherichia coli strain DH5 α was used throughout this study unless otherwise stated. BL-21(DE3) was used for radioactive labeling studies. Bacteria were propagated on Luria-Bertoni (LB) agar plates supplemented with ampicillin (100mg/ml) or spectinomycin (50 mg/ml) where required. For liquid cultures, bacteria were propagated in 2xYT for plasmid purification, PASM media for ^{35}S -methionine labeling and LB broth for protoporphyrin production assays.

PCR and uracil-excision cloning

Uracil-excision compatible PCR products were amplified with 28 cycles in 50 μl reaction mixtures using proof-reading PfuX7 polymerase³³. USER Fusion was performed as previously described²⁰ with minor modifications. Uracil-excision cloning cassettes were designed following the principles described by³. Uracil-excision compatible vectors were generated by PCR as described by³⁴ using the settings outlined above or by restriction digest using the enzymes Fast Digest AjuI (Thermo Scientific) and Nb.BtsI (New England Biolabs). Purified PCR products of inserts and vector were mixed in a total insert to vector ratio of 6:1 (volume). When more than one insert is cloned simultaneously, equimolar amounts of each insert was maintained. The reaction mixture was buffered with 5x Phusion HF buffer (Invitrogen) and 1 U of USERTM enzyme mix (New England Biolabs) were added. The reaction mixture was incubated for 20 min at 37 °C, followed by 20 min at 25 °C before transformation of chemically competent *E. coli* cells. All primers used in this study are summarized in Table S1

Plasmid construction

Compatible plasmid backbones were generated by inserting an NheI/PacI-digested AjuI-Nb.BsrDI cassette (GBlock, IDT-DNA, Table S1) into multiple cloning site-1 of pET-DUET (Novagen) using the XbaI and PacI restriction sites, generating pET-MTSG-Aju,BsrDI. The start codon for translation is contained within the 5'Nb.BsrDI site, while a stop codon is positioned downstream of the 3'Nb.BsrDI site. Entry clones of *hemB*, *E* and *F* were PCR amplified from *E. coli* K-12 genomic DNA with primer sets designed to eliminate internal TspRI sites. *Rhodobactor capsulatus hemA* and *E. coli hemC* and *D* were synthesized as TspRI-free genes flanked by the entry sequence and activated by TspRI-digest. All entry clones were introduced into a PCR amplified pET-MTSG-AjuI-BsrDI vector by uracil excision cloning as previously described³. Vector backbones for multi-gene assemblies were based on pCDF-DUET or pET-DUET (Novagen) and PCR amplified with uracil excision compatible primers. The primers were designed to introduce a *trc*-promoter and the A-linker upstream of the first insert as well as a T7-transcriptional terminator and the Db-linker downstream of the last insert. Entry clones for multi-gene assemblies were amplified with generic primers introducing a SD-sequence upstream and *ptrc* downstream. USER fusion assembly of entry clone inserts and backbones was essentially performed as previously described²⁰. Linkers were selected to match the backbone. All plasmids were validated by restriction digest patterns and sequencing. For a complete list of plasmids generated in this study consult Table S2.

³⁵S-labelling methionine labeling of proteins

Transcription and translation of protoporphyrin biosynthetic enzymes flanked by the 9bp entry sequence was confirmed by using the rifampicin blocking technique and ³⁵S-labeling³⁵. Pre-cultures of BL21(DE3) cells carrying relevant plasmids were sub-cultured (1:50) into 1 mL chemically defined rich medium lacking methionine³⁶. The cultures were grown at 37 °C to an optical density of 0.3-0.5, when expression was induced with 0.5 mM isopropyl-beta-D-1-thiogalactopyranoside. After 10 min of induction, the cells were treated with 25 mg/mL rifampicin for 10 min, labeled with 15 uCi/ml ³⁵S-methionine for 3 min, and harvested. The pellets were solubilized in 50 uL lysis buffer (CellLytic B, 1 mM EDTA, 2 mM MgCl₂, lysozyme, benzonase and complete protease inhibitor cocktail) and after 15 min incubation one volume of SDS-PAGE sample buffer was added (125 mM Tris-HCl pH 6.8, 20% glycerol, 4 % SDS,

100 mM DTT, Bromphenol blue). Samples were applied on precast 4-20% SDS-PAGE gels (Expedeon). Upon completion of the electrophoresis, the gels were dried and the labeled bands were visualized using a CyclonePus scanner (Perkin Elmer).

Screening for protoporphyrin production

Initial screening for protoporphyrin production was conducted directly from the plate used for transformation. Plates were incubated for 16 h at 37 °C after transformation to allow colony formation and then incubated for 72 h at 30 °C. At this stage, protoporphyrin production could be estimated by simple visual inspection due to red pigmentation accumulating in the colonies and the media.

Assay conditions and extraction procedure

For chromatographic assays, colonies were inoculated from plates into pre-cultures and grown 16 h at 37 °C and 300 rpm shaking. Assay cultures of 500ml fresh LB was inoculated with 10 ml pre-culture and grown in 2mL microtiter plates for 5 d at 30 °C, 300 rpm or 4 h at 37 °C, 300 rpm followed by 5 d at 16 °C, 300 rpm shaking. All screening assays were made in seven replicates. Extracts for screening were processed as follows: cells were precipitated and supernatant discarded, pellet was re-suspended in 200 ml H₂O and metabolites were extracted with 600 ml methanol at room temperature for 1 h. Cell debris was precipitated by centrifugation and cleared extract (600 ml) was acidified with 150 ml methanol pH 3.0, prior to analysis. For absolute quantification and compound identification we followed a procedure described by Kwon et al ²⁸ with minor modifications. Absolute quantifications were made in 3 or 8 replicates. Cells were grown in 2ml LB in 5ml microtiter plates for 5 d at 30 °C, 300 rpm. Protoporphyrins exported to the media were precipitated with 4 mg/ml DEAE Sephadex A-25. Following precipitation, pellets were extracted twice with 2ml Acetone:Acetonitrile:Acetic Acid (10:1:0.1) at 40 °C with sonication. Extracts were cleared by centrifugation, evaporated to dryness and re-suspended in 1 ml methanol.

Analytical method

Chromatographic separation of porphyrins was achieved using a five-minute isocratic method. Methanol adjusted to pH 3.0 with acetic acid was used as mobile phase and the column was an Ascentis Express Phenyl-Hexyl HPLC column (Supelco).

Chromatographic separation was achieved on a Dionex UltiMate 3000 UHPLC. Flow 0.4 ml/min, column oven at 30 °C. Fluorescence detection was made with a FLD-3000 Fluorescence Detector (Thermo Fisher Scientific, San Jose, CA) and the data processed by the Chromeleon 7.1 Chromatography Data System (Thermo Fisher Scientific). Mass spectra were recorded by an Orbitrap Fusion Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA) and data were processed with Xcalibur 3.0. Separation was achieved using the following parameters: Fluorescence detection parameters were 410nm for excitation and 620 nm emission. Mass spectrometry was performed in positive mode using electrospray ionization and full scan mode (350-2000 m/z). The injection volume was 20 µl for HPLC-FLD analysis, 10 µl for HPLC-FT-MS.

Supporting Information

Schematic examples of different backbone preparations, Fourier transform mass spectrometry (FT-MS) analysis of protoporphyrins and restriction analysis of selected stable and unstable clones. Material is available free of charge via the Internet at <http://pubs.acs.org>.

Acknowledgements

This work was funded by the Novo Nordisk Foundation

3.5 References

- (1) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 6, 343–345.
- (2) Walhout, A. J. M., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., van den Heuvel, S., and Vidal, M. (1999) GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. *Meth. Enzymol.* 328, 575–592.
- (3) Nour-Eldin, H. H., Hansen, B. G., Nørholm, M. H. H., Jensen, J. K., and Halkier, B. A. (2006) Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* 34, e122–e122.
- (4) Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2008) Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE* 4, e5553.
- (5) Blanus, M., Schenk, A., Sadeghi, H., Marienhagen, J., and Schwaneberg, U. (2010) Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. *Analytical Biochemistry* 406, 141–146.
- (6) Nielsen, J., and Keasling, J. D. (2011) Synergies between synthetic biology and metabolic engineering. *Nat. Biotechnol.* 29, 693–695.
- (7) Shetty, R., Lizarazo, M., Rettberg, R., and Knight, T. F. (2011) Chapter 13 - Assembly of BioBrick Standard Biological Parts Using Three Antibiotic Assembly. *Synthetic Biology Part B* 1st ed., pp 311–326. Elsevier Inc.

- (8) Heinemann, I. U., Jahn, M., and Jahn, D. (2008) The biochemistry of heme biosynthesis. *Archives of Biochemistry and Biophysics* 474, 238–251.
- (9) Tovmasyan, A., Sheng, H., Weitner, T., Arulpragasam, A., Lu, M., Warner, D. S., Vujaskovic, Z., Spasojevic, I., and Batinic-Haberle, I. (2013) Design, Mechanism of Action, Bioavailability and Therapeutic Effects of Mn Protoporphyrin-Based Redox Modulators. *Med Princ Pract* 22, 103–130.
- (10) Neuberger, A., and Scott, J. J. (1952) Aminolevulinic acid and Protoporphyrin Biosynthesis. *Nature* 172, 1093–1094.
- (11) Shemin, D., and Russell, C. S. (1953) δ -Amino Levulinic Acid, Its Role in the Biosynthesis of Protoporphyrins and Purines. *Journal of the American Chemical Society* 75, 4873–4874.
- (12) Beale, S. I., Gough, S. P., and Granick, S. (1974) Biosynthesis of delta-aminolevulinic acid from the intact carbon skeleton of glutamic acid in greening barley. *Proceedings of the National Academy of Sciences* 72, 2719–2723.
- (13) Beale, S. I., and Castelfranco, P. A. (1972) ^{14}C Incorporation From Exogenous Compounds into δ -Amino Levulinic Acid by Greening Cucumber Cotyledons. *Biochem Biophys Res Com* 52, 143–149.
- (14) Kannangara, C. G., Gough, S. P., Oliver, R. P., and Rasmussen, S. K. (1984) Biosynthesis of Δ -Aminolevulinate in Greening Barley Leaves VI. Activation of Glutamate by Ligation to RNA. *Carlsberg Research Communications* 49, 417–437.
- (15) Jahn, D., Michelsen, U., and Söll, D. (1990) Two glutamyl-tRNA reductase

activities in *Escherichia coli*. *Journal of Biological Chemistry* 266, 2542–2548.

(16) Neidle, E. L., and S, K. (1992) Expression of the *Rhodobacter sphaeroides* hemA and hemT genes, encoding two 5-aminolevulinic acid synthase isozymes. *Journal of Bacteriology* 175, 2292–2303.

(17) van der Werf, M. J., and Zeikus, J. G. (1995) 5-Aminolevulinate production by *Escherichia coli* containing the *Rhodobacter sphaeroides* hemA gene. *Applied and Environmental Microbiology* 62, 3560–3566.

(18) Kang, Z., Zhang, J., Zhou, J., Qi, Q., Du, G., and Chen, J. (2012) Recent advances in microbial production of δ -aminolevulinic acid and vitamin B12. *Biotechnology Advances* 30, 1533–1542.

(19) Bitinaite, J., and Nichols, N. M. (2009) DNA cloning and engineering by uracil excision. *Curr Protoc Mol Biol* (Bitinaite, J., and Nichols, N. M., Eds.) Chapter 3, Unit 3 21.

(20) Geu-Flores, F., Nour-Eldin, H. H., Nielsen, M. T., and Halkier, B. A. (2006) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res.* 35, e55.

(21) Tillett, D., and Neilan, B. A. (1999) Enzyme-free cloning: a rapid method to clone PCR products independent of vector restriction enzyme sites. *Nucleic Acids Res.* {27}.

(22) Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2010) A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE* (Peccoud, J., Ed.) 6, e16765.

- (23) Engler, C., Kandzia, R., and Marillonnet, S. (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE* (El-Shemy, H. A., Ed.) 3, e3647.
- (24) Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* 18, 6069–6074.
- (25) de Kok, S., Stanton, L. H., Slaby, T., Durot, M., Holmes, V. F., Patel, K. G., Platt, D., Shapland, E. B., Serber, Z., Dean, J., Newman, J. D., and Chandran, S. S. (2014) Rapid and Reliable DNA Assembly via Ligase Cycling Reaction. *ACS Synth. Biol.* 3, 97–106.
- (26) Anderson, J. C., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. A., Arkin, A. P., and Keasling, J. D. (2009) BglBricks: A flexible standard for biological part assembly. *J Biol Eng* 4, 1–1.
- (27) Nevin, D. E., and Pratt, J. M. (1990) A coupled in vitro transcription-translation system for the exclusive synthesis of polypeptides from the T7-promoter. *FEBS Lett.* 291, 259–263.
- (28) Kwon, S. J. S., de Boer, A. L. A., Schmidt-Dannert, C. C., 4. (2003) High-level production of protoporphyrins in metabolically engineered *Escherichia coli*: systematic extension of a pathway assembled from overexpressed genes involved in heme biosynthesis. *Appl Environ Microbiol* 69, 4875–4883.
- (29) Zelcbuch, L., Antonovsky, N., Bar-Even, A., Levin-Karp, A., Barenholz, U., Dayagi, M., Liebermeister, W., Flamholz, A., Noor, E., Amram, S., Brandis, A., Bareia, T., Yofe, I., Jubran, H., and Milo, R. (2012) Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* 41,

e98–e98.

- (30) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946–U112.
- (31) Salis, H. M. (2011) The Ribosome Binding Site Calculator. *Meth. Enzymol.* 498, 19–42.
- (32) Lee, M. J. (2012) Protoporphyrin Derivatives from a Recombinant *Escherichia coli* Grown on Chemically Defined Medium. *J. Microbiol. Biotechnol.* 22, 1653–1658.
- (33) Nørholm, M. H. H. (2010) A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol.* 10, 21.
- (34) Hansen, B. G., Mnich, E., Nielsen, K. F., Nielsen, J. B., Nielsen, M. T., Mortensen, U. H., Larsen, T. O., and Patil, K. R. (2012) Involvement of a Natural Fusion of a Cytochrome P450 and a Hydrolase in Mycophenolic Acid Biosynthesis. *Appl Environ Microbiol* {78}, {4908–4913}.
- (35) Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1989) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol* 185, 60–89.
- (36) Ukkonen, K., Vasala, A., Ojamo, H., and Neubauer, P. (2011) High-yield production of biologically active recombinant protein in shake flask culture by combination of enzyme-based glucose delivery and increased oxygen transfer. *Microbial Cell Factories* 10, 107.

Chapter 4

De-bugging and maximizing plant cytochrome P450 production in *Escherichia coli* with a scalable GFP-based optimization scheme

Cytochromes P450 are attractive enzyme targets in biotechnology as they catalyze stereospecific C-hydroxylations of complex core skeletons at positions that typically are difficult to access by chemical synthesis. Membrane bound CYPs are involved in nearly all plant pathways leading to the formation of high value compounds. In the present study, we systematically optimize the heterologous expression of six different plant-derived CYP genes in *Escherichia coli*, using a high-throughput workflow based on C-terminal fusions to the green fluorescent protein. The six genes can be over-expressed in a variety of *E. coli* strains using standard growth media.

Furthermore, sequences encoding a small synthetic peptide and a small membrane anchor markedly enhance the expression of all six genes. For one of the CYPs, the length of the linker region between the predicted N-terminal transmembrane segment and the soluble domain is varied, verifying the importance of this region for enzymatic activity. The described work describes how membrane bound CYPs are optimally produced in *E. coli* and thus adds this plant multi-membered key enzyme family to the toolbox for bacterial cell factory design.

4.1 Introduction

Escherichia coli is a model organism and a popular workhorse for protein production, with a comprehensive set of available genetic tools. An obvious extension to producing individual proteins for downstream applications is to produce a set of enzymes that catalyze an entire biosynthetic pathway. In our desire to advance from a petrochemical based society towards a bio-based society, it is critical that we deepen our understanding of how such heterologous biosynthetic pathways should be designed and maintained for the production of high value chemicals from renewable feedstock (Møller, 2014). Plant cytochrome P450 enzymes (CYPs) are central to this approach, as they are involved in nearly all pathways that lead to the formation of high value compounds such as terpenoids, alkaloids and phenylpropanoids (e.g. flavonoids, isoflavonoids, chalcones, aurones and lignans). Many of these compounds are used as medicines, condiments, flavors and fragrances (Morant et al., 2003), with well known examples being the antimalarial diterpenoid artemisinin (Chang et al., 2007), the anti-cancer drug paclitaxel (Chau et al., 2004), the flavor compound vanillin (Gallage and Møller, 2015; Hansen et al., 2009) and the steviol glucoside based-sweeteners (Davies and Deroles, 2014). Typically, the membrane anchored CYPs catalyze stereospecific hydroxylations of complex core carbon skeletons, at positions that are difficult to access by *de novo* chemical synthesis. CYPs often show high substrate specificity and although members of this ancient multigene family are found in all domains of life, plants seem to be particularly enriched with often more than 250 CYPs in a single species (Nelson, Biochim Biophys Acta (2011) 1814: 14-18). Unfortunately, plant CYPs have the reputation of being notoriously difficult to produce in *E. coli* (Chang et al., 2007). Furthermore, in order to work efficiently, CYPs often need co-production of other membrane-associated proteins such as reductases (Eugster et al., 1992) or cytochromes b5 (Paddon et al., 2013; Zhang et al., 2007).

Previous approaches to improve plant CYP production in *E. coli* includes exchanging parts of the native amino-terminal region with parts of the amino-terminal region of a codon optimized bovine CYP17alpha; often referred to as the “Barnes sequence” (Bak et al., 1998a; 1997; Barnes et al., 1991; Leonard and Koffas, 2007). Similarly, bacterial expression of human CYPs have been aided by bacterial leader

sequences such as the membrane translocation signals from *pelB* and *ompA* (Pritchard et al., 1997); for a recent review see (Zelasko et al., 2013)

Membrane protein production in *E. coli* has benefited greatly from the introduction of a streamlined green fluorescent protein-based pipeline for rapid and simple assessment of proper expression (Drew et al., 2006; 2005; 2001; Lee et al., 2014). Briefly, a carboxy-terminal GFP-fusion has many advantages: without interfering with membrane targeting and integration, it acts as an expression reporter as fluorescence indicates that the mRNA was translated in full. Importantly, if the upstream protein is mislocalized and subsequently aggregated, GFP fluorescence is completely quenched. Further, once folded, GFP remains fluorescent even during otherwise denaturing applications such as SDS-PAGE. In summary, the carboxy-terminal GFP provides a cheap, reliable and fast readout of the proper production of membrane proteins. Moreover, the fluorescent properties of such chimeric proteins can be used for screening the optimal solubilization conditions for downstream applications, such as structural studies (Kawate and Gouaux, 2006; Sonoda et al., 2011).

Currently, there is no established formula for the heterologous expression of CYPs in *E. coli*. The literature is scattered with a wealth of different genetic manipulations and use of various strains and growth conditions. To enable efficient heterologous CYP production, there is an obvious need to systematically determine proper and generally useful expression conditions. Here, we demonstrate how functional expression of six different plant CYPs in *E. coli* is achieved by very simple means. The carboxy-terminal GFP-approach readily lends itself to high throughput applications, and therefore the proposed setup is bound to be highly useful for large scale expression screens.

4.2 Materials and Methods

Bacterial strains

Escherichia coli strain NEB 5-alpha (New England BioLabs, Ipswich, USA) was used for cloning of PCR products and propagation of plasmids. The following *E. coli* strains were used for gene expression: Rosetta2(DE3) pLysS (Novagen, Merck Millipore, Germany), B121(DE3) pLysS (Promega, Madison, USA), C41(DE3) (Miroux and Walker, 1996), KRX (Promega, Madison, USA), MC4100(DE3)pLysS and MG1655(DE3)pLysS.

PCR and uracil excision

All DNA manipulations were performed using uracil excision technology as previously described (Nour-Eldin et al., 2006; Nørholm, 2010) and all oligonucleotides are listed in Table S1. PCR products were amplified in 50 μ L reactions containing: 1 μ L PfuX7 DNA polymerase, 0.2 mM dNTPs (Thermo Scientific, Waltham, USA), 1.5 mM $MgCl_2$, 0.5 μ M forward oligonucleotide, 0.5 μ M reverse oligonucleotide, Phusion® HF Reaction Buffer (New England BioLabs, Ipswich, USA) and 50 ng plasmid template. A touch-down PCR program was used for amplification: Step 1: 2 min 98°C; step 2: 15 sec 98°C, 20 sec 65°C (-1°C per cycle), 45 sec per kb at 72°C (step 2 repeated 9 times until 55°C, then repeated 20 cycles at the annealing temperature 55°C); step 3: 5 min 72°C; step 4 hold at 10°C. PCR products were gel purified from 1% (w/V) agarose gel using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Düren, Germany) and eluted in 10% TE buffer. Purified PCR products were incubated with 1 μ L USER™ enzyme (New England BioLabs, Ipswich, USA) for 30 min at 37°C and subsequently mixed with linearized vector backbone in a molar ratio 3:1 and incubated at 18°C for 1 h. A Nanodrop spectrophotometer 2000 (Thermo Scientific, Waltham, USA) was used for estimation of PCR product and vector concentration. Approx. 5 μ L of the assembled PCR product:vector solution was transformed into NEB 5-alpha chemically competent cells according to the manufacturer's protocol. Transformants were selected on Luria Bertoni (LB) agar plates supplemented with 50 μ g/mL kanamycin, 25 μ g/mL chloramphenicol or 10 μ g/mL gentamycin. Colonies were screened for gene insert by colony PCR using OneTaq 2X Master mix (New England BioLabs, Ipswich, USA).

Vectors were extracted and purified using QIAprep Spin Kit (Qiagen) and verified by DNA sequencing (Eurofins Genomics, Ebersberg, Germany).

DNA constructs

All DNA constructs were made with uracil excision as previously described (Nour-Eldin et al., 2006; Nørholm, 2010). Details of oligonucleotides, template DNA and references can be found in Tables S1 and S2 and below. Briefly, *AsiSI* restriction enzyme recognition sites, present in the kanamycin resistance cassettes, were deleted from all pET28a(+)-derived construct . *AsiSI* uracil excision compatible cloning cassettes were inserted between sequences encoding a TEV protease site, GFP and polyhistidine tag and either the 28-tag (Nørholm et al., 2013) (Fig. 1A), a Barnes-like N-terminal sequence MALLLAVF, SohB (residues 1-48, KDT39511) or YafU (residues 1-88, KEN61237) . Six genes encoding different plant CYPs (*Sorghum bicolor* CYP51G1 (U74319); *Avena strigosa* CYP51H10 (DQ680849); *Sorghum bicolor* CYP71E1 (O48958); *Sorghum bicolor* CYP79A1 (U32624); *Arabidopsis thaliana* CYP79B2 (NM_120158) and *Picea sitchensis* CYP720B4 (HM245403)) were PCR amplified with sequence specific oligonucleotides and cloned into the different pET28-derived vectors either treated with the restriction enzyme *AsiSI* (Thermo Scientific, Waltham, USA) and the nicking enzyme *Nb.BbvCI* (New England Biolabs, Ipswich, USA) or PCR amplified. The *sohB*- and *yafU*-based *E. coli* membrane anchors have their C-termini on the cytoplasmic side of the *E. coli* inner membrane (Daley et al., 2005) and therefore the plant CYPs cloned into the corresponding backbones had their predicted N-terminal anchors truncated to avoid translocation of the CYP into the periplasmic space. Truncations were designed using TMHMM v. 2.0 (Sonnhammer et al., 1998). A Strep-HRV3C tagged, codon optimized *Sorghum bicolor* CPR2b (Wadsäter et al., 2012) was cloned into pET28a(+)-*tev-gfp-his8* and *gfp* was subsequently deleted and the origin of replication was swapped with the corresponding parts from pSEVA63 (Silva-Rocha et al., 2013) by amplifying pSEVA63 with the oligo nucleotides 5'-ATCCGCTUTAATTAAAGGCATCAAATAAAAC-3' and 5'-ACTAGTCTUGGACTCCTGTTGATAGATC-3' and the pET28-based *cpr2b* construct with the oligo nucleotides 5'-AAGCGGAUCTACGAGTTGCATGATAAAGAAGACAGTC-3' and 5'-AAGACTAGUCAATCCGGATATAGTTCCTCCTTTTCAG-3'. A truncated version

of *E. coli lepB* was amplified from a previously described pGem1-Lep construct (Hessa et al., 2005) using the oligonucleotides 5'-ACTCGAGGAUGGCGAATATGTTTGGCCCTGATTC-3' and 5'-ATCGCTGCUTCCAGGACCACCACTAGTCTCG-3' and combined with the pET28-based full length *CYP* constructs amplified with 5'-ATCCTCGAGUCTCCTTCTTAAAG-3' in combination with the gene specific forward oligo nucleotides.

Culture media and expression conditions

All strains were grown aerobically in liquid cultures. For plasmid propagation single colonies were grown over night in 2xYeast/Tryptone medium at 37°C 250 rpm. For single plasmid transformations, chemically competent cells were transformed with 20 ng plasmid according to the manufacturer's protocol. For co-transformation of plasmids for expression assays, electrocompetent cells were transformed with 20 ng of plasmid DNA. Four different media were tested for expression, LB (1% tryptone; 0.5% yeast extract; 1% NaCl), Terrific Broth (TB) (1.2% tryptone; 2.4% yeast extract; 0.4% glycerol; 17 mM potassium phosphate (monobasic); 72 mM potassium phosphate (dibasic)); the defined rich medium PA-5052 (Studier, 2005) (50 mM Na₂HPO₄; 50 mM KH₂PO₄; 25 mM (NH₄)₂SO₄; 2 mM MgSO₄; 10 µM metals; 0.5% glycerol; 0.05% glucose; 0.2% alpha-lactose; 200 µg/mL amino acids E, D, K, R, H, A, P, G, T, S, Q, N, V, L, I, F, W and M). and minimal M9 (M9 salts; 2mM MgSO₄; 0.1mM CaCl₂; 0.2% glycerol; 0.2% glucose; 10 µM Fe³⁺; 200 µg/mL amino acids E, D, K, R, H, A, P, G, T, S, Q, N, V, L, I, F, W and M). For media tests, over night cultures were prepared with each of the four media supplemented with 0.5% (w/V) D-glucose and appropriate antibiotics. The pre-cultures were subsequently inoculated into each of the corresponding media for expression. For other expression assays, over night cultures were made in 96 deep well plates, inoculating single colonies into 800 µL TB supplemented with 0.5% (w/V) D-glucose and appropriate antibiotics and grown at 30°C 250 rpm in Innova®44R incubator shaker system (5 cm orbital shaking) (New Brunswick Scientific, Eppendorf, USA). The optical density (OD) of the over night cultures was measured at Abs_{600nm} on Plate Reader SynergyMx (SMATLD) (BioTek, Winooski, USA). Over night cultures were inoculated into 5

Chapter 4

mL fresh TB medium in 24 deep well plates to a final OD of 0.05. Cells were incubated for approx. 2 h at 37°C 250 rpm to an OD of 0.3-0.5. All strains were induced with a final conc. of 0.4 mM isopropyl β -D-1-thiogalactopyranoside (IPTG, dioxane free, Thermo Scientific, Waltham, USA). Despite the autoinduction capacity of the PA-5052 medium, as described previously (Lee et al., 2014), IPTG was used for induction in this medium as well. The strain KRX was furthermore induced with a final conc. of 5 mM L-rhamnose (Sigma-Aldrich, St. Louis, USA). Cultures were subsequently incubated at 25°C 150 rpm for 3 h or 22 h.

Whole cell fluorescence measurements

Whole cell fluorescence was measured using 2 mL induced culture. The cells were harvested (2,500xg, 20 min) and resuspended in a total of 100 μ L PBS buffer. The GFP fluorophore was allowed to form for 1 h at room temperature and then fluorescence was detected using excitation at 485nm and emission at 512nm with a window of \pm 9 nm, gain value 50, using plate reader SynergyMx SMATLD (BioTek, Winooski, USA).

SDS-PAGE

Whole cells were lysed for 1.5 h at room temperature in TRIS buffer (50 mM Tris HCl pH 7.5; 150 mM NaCl; 2 mM MgCl_2) with 250 U/mL Benzonase[®] nuclease (Sigma-Aldrich, St. Louis, USA); 5 mg/mL lysozyme egg white powder (Amresco, US, Ohio) and the cOmplete ULTRA EDTA-free protease inhibitor cocktail (Roche, Basel, Switzerland). The cell lysate ($\text{OD}_{600\text{nm}}$ 0.1) was analyzed in parallel with PageRuler[™] Prestained Protein Ladder 10-170K (Thermo Scientific, Waltham, USA) by standard SDS-page using Mini-PROTEAN[®] TGX[™] 4-15% gels (Bio-Rad, Hercules, USA). In-gel fluorescence was detected on a G:BOX UV-table (Syngene, Cambridge, UK).

Enzyme activity measurements

For CYP79A1 enzymatic assays, cells were harvested by centrifugation 2,500xg 4°C for 10 min, washed once in 50 mM Potassium phosphate (KPi) buffer pH 7.5 and

resuspended to 0.03 OD units per μL in 50 mM KPi buffer. The enzyme reaction was carried out in 30 μL volume consisting of 5 mM NADPH, 0.5 mM L-Tyrosine (Sigma-Aldrich, St. Louis, USA), 50 mM KPi buffer and 20 μL cell suspension. Cells were incubated at 30°C 400 rpm for 60 min. The product (*E*)-*p*-hydroxyphenylacetaldoxime (oxime) was extracted by adding 150 μL methanol followed by incubation at room temperature for 10 min. Cells debris was discarded twice by centrifugation (20,000 \times g, 10 min) and the supernatant was transferred into HPLC vials and stored at -20°C prior to analysis. The oxime was also extracted directly from cultures co-expressing *SbCYP79A1* and *SbCPR2b*. 100 μL culture was extracted with 100 μL methanol, as described above.

LCMS detection

The CYP79A1 product (*E*)-*p*-hydroxyphenylacetaldoxime (Mw 151.17) was detected and quantified by LC-MS. A chemically synthesized (*Z*)-*p*-hydroxyphenylacetaldoxime standard was kindly provided by Mohammed Saddik Motawie (University of Copenhagen, Department of Plant and Environmental Sciences). The two geometrical isomers (*Z*)-*p*-hydroxyphenylacetaldoxime and (*E*)-*p*-hydroxyphenylacetaldoxime were detected in both the chemical oxime standard and the oxime sample product due to instability and chemical equilibrium. Thus for the purpose of total oxime quantification the chromatogram area of the both peaks were summed. LC-MS data was collected on a Bruker Evoq triple quadrupole mass spectrometer equipped with an Advance UHPLC pumping system. Samples were held in the CTC PAL autosampler at a temperature of 10.0 °C during the analysis. Injections (2 μL) of the sample were made onto a Supelco Discovery HS F5-3 HPLC column (3 μm particle size, 2.1 mm i.d. and 150 mm long). The column was held at a temperature of 30.0 °C. The solvent system (flow rate: 1.0 ml/min) used was water with (A) 100mM ammonium formate and (B) acetonitrile using the following elution profile: 0.5 min 95% A/5% B, linear gradient to 50% A/50% B for 3.0 min, 1.5 min 50% A/50% B and re-equilibration for 2 min 95% A/5% B. The column eluent flowed directly into the heated ESI probe of the MS, which was held at 350°C and a voltage of 4500 V. SRM data was collected in centroid at unit mass resolution. Positive ion mode with Q1 set to monitor 152.70 m/z , Q3 set to monitor 136 m/z and Q2 set to a collision energy of 10.0eV, with an Argon pressure of 1.5 mTorr. The other MS

settings were as follows, Sheath Gas Flow Rate of 40 units, Aux Gas Flow Rate of 40 units, Sweep Gas Flow Rate of 20 units, Ion Transfer Tube Temp was 350 °C.

4.3 Results

Full length expression of six plant CYPs in E. coli grown in standard media

Previously, it was shown that simple addition of a 84 nucleotide/28 amino acid tag (28-tag) to the amino-terminus greatly enhanced the expression of challenging membrane protein encoding genes in *E. coli* (Nørholm et al., 2013). Here, we used a similar design principle to sandwich plant CYPs between an amino-terminal 28-tag and a carboxyterminal GFP (Fig. 1A). Six CYP genes from four different mono- and dicotyledonous plants were cloned into the pET28-28tag-tev-gfp-his8 plasmid: *Sorghum bicolor* CYP51G1 (Bak et al., 1997), CYP71E1 (Bak et al., 1998a) and CYP79A1 (Koch et al., 1995), *Avena strigosa* CYP51H10 (Qi et al., 2006), *Arabidopsis thaliana* CYP79B2 (Bak et al., 1998b); and *Picea sitchensis* CYP720B4 (Hamberger et al., 2011) and transformed into *E. coli* KRX. Following three hours induction in four different standard growth media (LB, TB, M9 supplemented with Fe³⁺ and the synthetic rich medium PA-5052); expression and protein integrity was analyzed by measuring whole cell- and in-gel fluorescence. Compared to an empty vector control, all six constructs exhibited significantly elevated fluorescence levels in all growth media except M9 (Fig. 1B). Moreover, analysis of the samples by in-gel fluorescence indicated that the majority of the constructs were expressed as full-length proteins (Fig. 1C), except for CYP51H10 for which a significant proportion seemed to be truncated or partly degraded. Addition of 5-aminolevulinic acid, suggested to aid the functional production of heme-containing proteins (Sudhamsu et al., 2010) did not have any apparent effect on expression levels (data not shown). Based on the highest expression level and cell density, we chose TB medium for further expression optimization.

Chapter 4

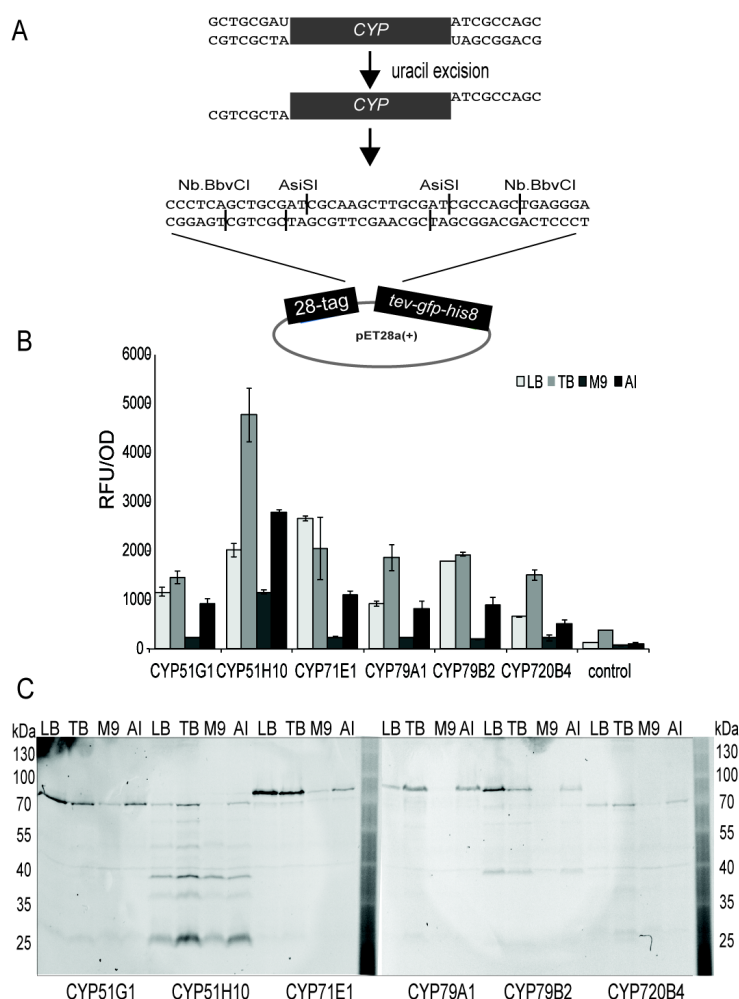


Figure 1. Production of six different CYP enzymes in standard growth media. (A) Schematic representation of the cloning procedure used. Genes encoding six different CYPs were PCR amplified with specific uracil-containing tailed oligonucleotides and assembled with a compatible *AsiSI/Nb. NbvCI* treated vector. **(B)** Effect of media on production of CYP51G1, CYP51H10, CYP71E1, CYP79A1, CYP79B2 and CYP720B4 in the KRX strain. Four media types were tested: Luria Bertoni (LB), Terrific Broth (TB), minimal media M9 and autoinducing media (AI). As cell cultures reached the exponential growth phase, cultures were induced for 3 h. Cell cultures were subsequently harvested by centrifugation and resuspended in PBS buffer for whole cell fluorescence detection (relative fluorescence unit (RFU) per optical density (OD)). Error bars indicate standard error of the mean (n=3). Empty vector (control) was included as reference. **(C)** The level of full-length GFP-fused protein monitored from cell lysate by in-gel fluorescence detection.

E. coli strains KRX and Rosetta excel in plant CYP expression

Next we compared expression levels of the six different CYP constructs in the six different *E. coli* strains MC4100(DE3) pLysS, MG1655(DE3) pLysS, KRX, BL21(DE3) pLysS, Rosetta2(DE3) pLysS, and C41(DE3). In all strains except for KRX and C41, expression from the T7 promoter is catalyzed by the T7-phage derived RNA polymerase controlled by the PlacUV5 IPTG-inducible promoter. In three of the strains, potential toxic effects of over-expression is counteracted by expression of the natural T7 DNA polymerase inhibitor T7 lysozyme from the pLysS or the pRARE

plasmid (Studier, 1991), whereas a similar effect is obtained in C41 by mutations in PlacUV5 that make the promoter less strong (Wagner et al., 2008). In the commercial strain KRX expression of T7 RNA polymerase is tightly controlled by a rhamnose inducible promoter.

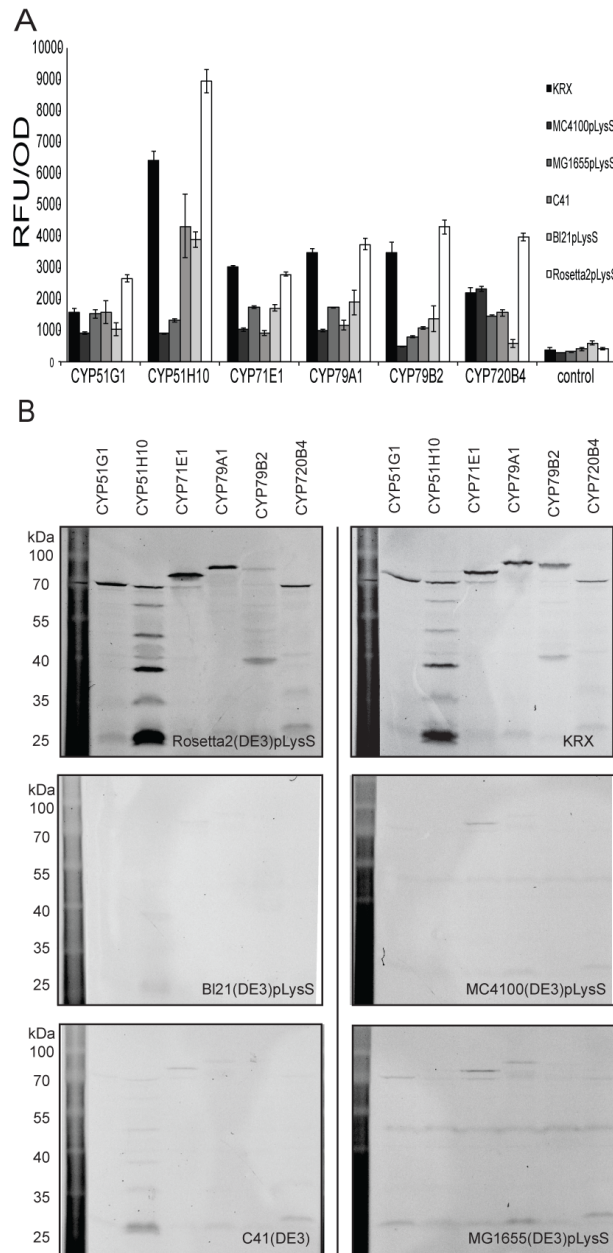


Figure 2. Strain-dependent production of CYP enzymes. (A) Three K-strains (KRX, MC4100pLysS, MG1655pLysS) and three B-strains (C41, BI21pLysS, Rosetta2pLysS) were transformed with 28-tag:CYP:tev-gfp-his8 expression constructs. Cells were grown in TB media and induced for 3 h, as cell cultures reached the exponential growing phase. Overall expression was monitored by whole cell fluorescence detection (relative fluorescence unit (RFU) per optical density (OD)). Error bars indicate standard error of the mean (n=3). Empty expression vector was included as control. **(B)** The level of full-length GFP-fused protein was monitored from cell lysate by in-gel fluorescence detection in the three B-strains (left panel) and the three K-strains (right panel).

The six pET28-28tag-*CYP-tev-gfp-his8* constructs were transformed into the six different strains and expression was monitored by whole cell- (Fig. 2A) and in-gel- (Fig. 2B) fluorescence after 3 h induction. As judged from whole cell fluorescence and presence of full-length protein on SDS-gels, strains KRX and Rosetta2(DE3) pLysS clearly outperformed the other four strains (Fig. 2A and 2B).

A transmembrane domain encoded by E. coli sohB normalizes expression of plant CYP genes to high levels

Most eukaryotic CYPs utilize N-terminal hydrophobic peptides to localize to the endoplasmic reticulum, and sequence modifications in this region have previously proven essential for heterologous expression (Barnes et al., 1991). We hypothesized that it would be beneficial to exchange the plant ER/membrane localization signals from the CYPs with membrane anchors that are derived from the heterologous expression host and that have previously been shown to express at high levels from the T7 promoter. To find suitable candidates, we mined a previously published library of membrane protein encoding genes expressed in *E. coli* (Daley et al., 2005). We chose three of the most highly expressed genes (Fig. 1S.) that encoded membrane proteins with different topologies: *sohB*, encoding a single-pass membrane protein with the C-terminal end localized inside the bacterial cytoplasm (C_{in}), *yafU*, encoding two membrane spanning regions with a C_{in} topology; and *lepB*, encoding two membrane spanning regions, but with the opposite (N_{out}) topology (Fig. 3A). To mimic the overall topology of the microsomal plant CYPs and ensure localization of the CYP catalytic domain to the cytoplasm, we replaced the hydrophobic sequences of the CYPs with the transmembrane parts of SohB and YafU; whereas the full length CYP was fused to the periplasmic carboxy-terminus of LepB. For benchmarking purposes, we added a Barnes-like MALLAVF-encoding sequence to the six *CYP* genes and comparisons were made to unmodified plant CYP sequences. The native controls, 28-tag-, MALLAVF-, *sohB*-, *yafU*- and *lepB*-tagged *CYP* constructs were transformed into KRX and Rosetta2(DE3) pLysS and expression was monitored by whole cell (Fig. S2) and in-gel fluorescence (Fig. 3B). None of the native N-terminally unmodified sequences expressed to any significant levels. In contrast, both the 28-tag and the SohB-domain normalized the expression of the *CYPs* to high levels and in particular the combination of the KRX strain with the SohB-tagged CYPs

consistently gave high levels of full-length protein. Furthermore, compared to the 28-tag constructs, a higher proportion of SohB-tagged CYP79A appeared as full-length, both 3 h and 22 h-post-induction (Fig. S3)

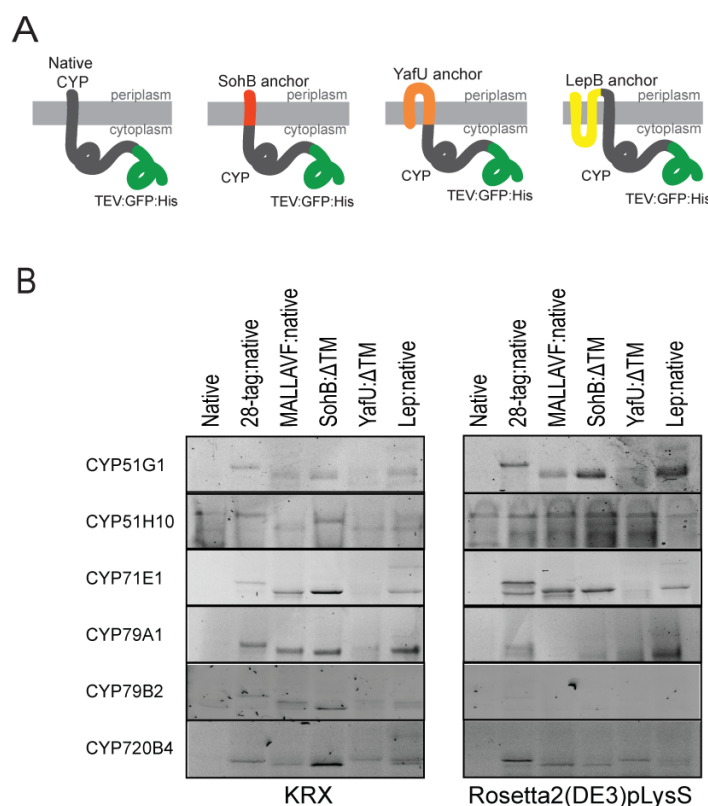


Figure 3. Effect of different N-terminal peptides and small bacterial membrane anchors on the expression of six different CYPs. (A) Illustration of the hypothetical membrane topology of native CYPs and recombinant bacterial anchor constructs: (from left) reference CYP with native N-terminal sequence, SohB membrane domain (1TM, highlighted in red) fused with a truncated CYP, YafU anchor (2TM, C_{in}, highlighted in orange) fused with a truncated CYP, and LepB anchor (2TM, C_{out}, highlighted in yellow) fused to a native CYP. All constructs have TEV:GFP:His8 fusions at their C-terminal. (B) Six CYPs were tested with different N-terminal modifications: (from left) native sequence as reference, 28-tag fused to native CYP sequence, Barnes-like MALLAVF peptide fused to native CYP sequence, SohB membrane domain fused to a truncated CYP sequence, YafU membrane domain fused to a truncated CYP sequence and the LepB membrane domain fused to a native sequence. Constructs were expressed in two strains (KRX and Rosetta2(DE3)pLysS). Cell cultures were induced for 3 h before harvest. Cell lysates were analysed by in-gel fluorescence to determine the level of full-length protein present.

Activity is preserved in all engineered versions of CYP79A1 except for those truncated in a proline rich region that precedes the soluble catalytic domain

To obtain detectable expression levels of CYPs, we substantially re-engineered plant CYP proteins with truncations and/or heterologous protein domains such as the five different amino-terminal peptides and the large carboxy-terminal GFP domain described above. To test that a CYP is able to retain activity with these major modifications, we functionally assayed all the modified versions of CYP79A1 both *in*

vivo and *in vitro* while monitoring expression levels by fluorescence. Moreover, we checked the effect of truncating CYP79A1 at the residue Y36, likely placed at the interface between the membrane and the aqueous face; at the residue P69, placed at a proline-rich, proposed hinge region observed in many P450s (Chen et al., 1998; Leonard and Koffas, 2007; Williams et al., 2000); and at two positions in between these two extremes (P44 and P60, Fig. 4A).

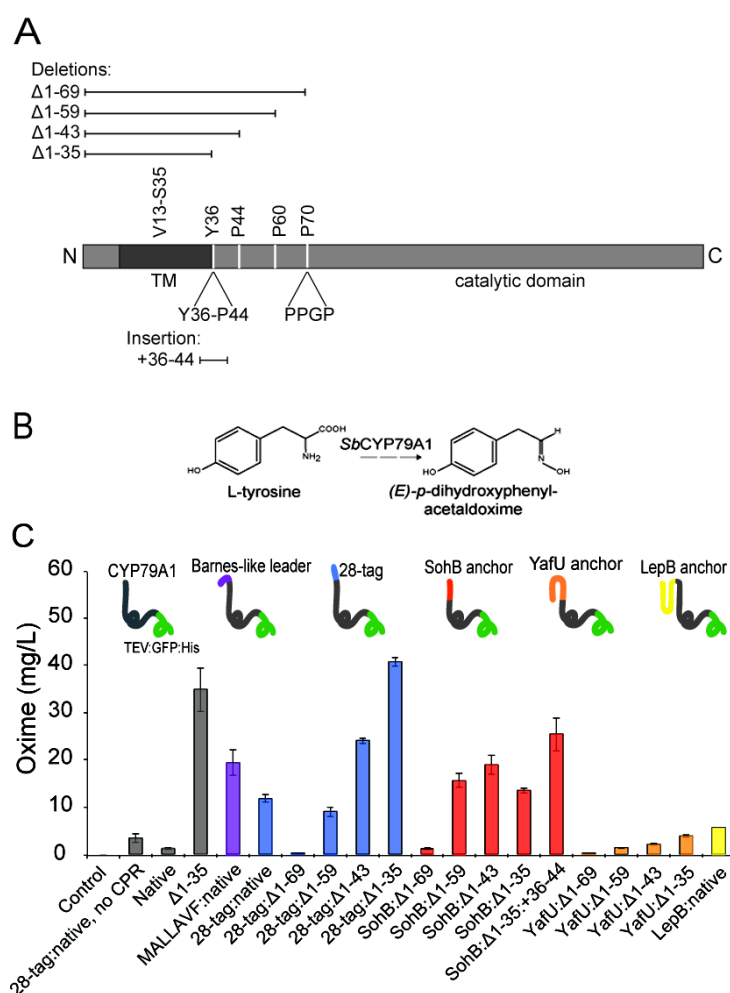


Figure 4. Functional assaying of engineered variants of CYP79A1. (A) Schematic overview of CYP79A1 modified expression constructs used in this study. Four truncated expression constructs were designed to remove various parts of the N-terminal sequence: $\Delta 1-35$: at the end of the predicted transmembrane segment between residues 1 and 35, and at three proline residues at positions 44, 60 and 70 ($\Delta 1-43$, $\Delta 1-59$, $\Delta 1-69$) preceding the catalytic domain. In addition, an artificially extended linker region was engineered by inserting a repeat of the amino acids Y36-P44. (B) CYP79A1-catalyzed conversion of L-Tyrosine to (E)-p-hydroxyphenylacetaldoxime. (C) Differently truncated and engineered CYP79A1 constructs were co-produced with the compatible reductase electron donor CPR2b in the KRX strain. Cells were induced for 22 h, before oxime was extracted from the culture. Empty vector control and 28-tag:native CYP79A1 expressed without CPR were included as references. Error bars indicate standard error of the mean (n=3).

In the *in vitro* assay, CYP79A1 and the reductase *Sorghum bicolor* CPR2b were co-expressed in the KRX strain and the activity was measured on harvested cells by

adding NADPH and the substrate tyrosine. Formation of the product (*E*)-*p*-hydroxyphenylacetaldoxime (Fig. 4B) was followed by separation using high performance liquid chromatography and mass detection (LCMS) (Fig. S4 and S5A). In parallel, we performed a quantitative *in vivo* assay by co-expressing *CYP79A1* with *SbCPR2b* and monitoring oxime formation with LCMS detection (Fig. 4C). Whole cell fluorescence confirmed expression of all constructs except the construct with the native, unmodified *CYP79A1* sequence, as described above (Fig. S5B), and most of the expressed constructs were catalytically active except for those with the most heavily truncated *CYP79A1* sequence at position P69. Cells expressing the *YafU* and *LepB* fusions were less active than those expressing the 28- and the *SohB*-fusions (Fig. 4B). In order to see if the distance to the membrane is important, we increased the linker between the membrane anchor and the soluble domain by duplicating the sequence Y36-P44 in the *SohB*-*CYP79A1* fusion construct. The artificially increased size of the linker resulted in a seemingly small increase in enzyme activity (Fig. 4B). Finally, particularly because the CYP needs to functionally interact with its reductase partner, we tested if the GFP fusion had a negative impact on the proper interactions, but only observed minor improvements in the activity when GFP was removed (Fig. S6).

4.4 Discussion

The multitude of available genetic tools, growth media, and genotypes for the model organism *E. coli* is a blessing, but can also be a curse. Optimization of the physical parameters for heterologous gene expression is a multifactorial problem, and the details - from transcription to translation and post-translational processing - can be endlessly tinkered with. To this end, a fast reliable and inexpensive assay to optimize expression conditions in a high-throughput fashion is of great importance, especially for the design of cell factories that are based on expression of multiple challenging enzymes like CYPs. Membrane bound CYPs are involved in nearly all plant pathways leading to the formation of high value compounds. GFP has been demonstrated to be a useful reporter of proper expression and protein folding for soluble proteins and membrane proteins alike (Drew et al., 2006; 2005; 2001; Lee et al., 2014; Waldo et al., 1999).

Here, we have explored the use of GFP fusions to report on the functional production of a commercially attractive group of single spanning membrane proteins, the multi-gene encoded cytochrome P450 enzyme family (Morant et al., 2003). We discovered that the two expression strains KRX and Rosetta2(DE3) pLysS consistently outperforms the other tested expression strains with respect to the amount of functionally active P450 enzyme produced under short term induction conditions. KRX is a K strain and Rosetta2 is a B strain, suggesting that there is no obvious restriction in the use of K vs. B strains for *CYP* expression. The Rosetta2 strain provides additional copies of tRNAs that are present in low concentration in *E. coli* strains, whereas KRX does not, suggesting that this complementation is not the major causative effect on *CYP* expression in Rosetta2. Rather, a likely similarity between the two strains is the carefully balanced expression of the T7 RNA polymerase – in KRX by controlling expression directly from the Prha promoter and in Rosetta2 indirectly by expression of *lysS* from the pRARE plasmid. Several of the other strains express *lysS* from the pLysS plasmid, but these do not perform as well, and we speculate that the difference in *lysS* carrying plasmid may change the production of lysozyme in a way that affects *CYP* expression. In line with this, upon removal of *lysS* from pLysS, we obtained similar expression levels in Rosetta and BL21 with all six P450s (data not shown). Our comparison of growth media suggests that there is no

“magic” medium for *CYP* expression, but that the minimal medium M9 without additional supplements is suboptimal.

Another finding is that expression levels can be normalized to high levels using small peptides such as the 28-tag or membrane-spanning domains like SohB. Similar effects have been observed before with tags such as polyhistidine, the maltose binding protein (MBP) and the small ubiquitin related modifier (SUMO, for a recent review see (Costa et al., 2014)). The frequently observed positive effect of adding sequences to the 5'-end is possibly due to an incompatibility between the native plant 5' sequences with high-level expression, as previously reported for membrane protein encoding genes like *araH* and *narK* (Nørholm et al., 2013). An alternative solution that previously has proven successful is to introduce synonymous changes in the first couple of codons downstream from the start codon. Indeed, we have successfully expressed the native *CYP79A1* and *CYP71E1* genes by introducing a few codon changes in the pET expression system (data not shown). We hypothesized that the inherent membrane targeting of e.g. SohB would have a positive impact of expression *and* activity of an ER-derived protein targeted to the *E. coli* inner membrane, but found no significant enhancement in activity when comparing SohB with e.g. 28-tagged CYPs. However, we also did not observe any major changes in activity (Fig.4b) or membrane localization of CYP79A1 even when completely removing all N-terminal (predicted) transmembrane segments (data not shown). Membrane association in the absence of N-terminal hydrophobic sequence has been observed before (Doray et al., 2001), suggesting that some CYPs are sufficiently hydrophobic to localize to the membrane in the absence of an N-terminal localization signal (Jensen et al., 2011). This is also supported by molecular dynamics studies suggesting interactions between helices located far downstream from the N-terminus and biological membranes (Denisov et al., 2012). The identification of SohB as a generic tool to enhance expression, while being native to the *E. coli* membrane insertional machinery, may become useful when engineering higher order membrane assemblies such as the suggested multi-CYP metabolons (Laursen et al., 2015; Moller, 2010). Also, even though the *lepB* fusions expressed to a lower level than those based on e.g. SohB, the chimeric proteins were active and may become useful because they maintain the native plant membrane anchor and (presumably) topology. This is particularly relevant because our results clearly demonstrate that care should

be taken when truncating CYPs, as exemplified with the inactive P69-truncated CYP79A1.

4.5 Conclusion

The presented work demonstrates the usefulness of the GFP-reporter approach for de-bugging and maximizing expression of a group of related enzymes, and suggests that there are no inherent limitations in using different standard *E. coli* strains and expression conditions for exploiting CYPs for biotechnological applications. Further, the peptides and membrane domains used in this study add new biobricks to a toolbox for engineering pathways of higher complexity such as the taxol biosynthetic pathway consisting of eight steps catalyzed by CYPs (Chau et al., 2004).

Acknowledgements

We thank Victor de Lorenzo and the members of his laboratory for generously providing the pSEVA collection. We thank Tomas Laursen, Peter Naur, Søren Bak, Björn Hamberger, Johan Andersen-Ranberg and Britta Hamberger for advice on CYPs and reductases. We thank David Drew, Daniel Daley and Jan Willem de Gier for discussions on *E. coli* gene expression and the GFP-based expression platform. SS is the recipient of VILLUM Foundation's Young Investigator Programme grant VKR023128. This work was supported by the Novo Nordisk Foundation, from the VILLUM research center of excellence “Plant Plasticity”, from the UCPH Excellence Program for Interdisciplinary Research to Center of Synthetic Biology ”bioSYNergy” and by an European Research Council Advanced Grant Project No. 323034: LightdrivenP450s.

4.6 References

- Bak, S., Kahn, R.A., Nielsen, H.L., Moller, B.L., Halkier, B.A., 1998a. Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. *Plant Mol. Biol.* 36, 393–405.
- Bak, S., Kahn, R.A., Olsen, C.E., Halkier, B.A., 1997. Cloning and expression in *Escherichia coli* of the obtusifolius 14 α -demethylase of *Sorghum bicolor* (L.) Moench, a cytochrome P450 orthologous to the sterol 14 α -demethylases (CYP51) from fungi and mammals. *Plant J.* 11, 191–201.
- Bak, S., Nielsen, H.L., Halkier, B.A., 1998b. The presence of CYP79 homologues in glucosinolate-producing plants shows evolutionary conservation of the enzymes in the conversion of amino acid to aldoxime in the biosynthesis of cyanogenic glucosides and glucosinolates. *Plant Mol. Biol.* 38, 725–734.
- Barnes, H.J., Arlotto, M.P., Waterman, M.R., 1991. Expression and enzymatic activity of recombinant cytochrome P450 17 α -hydroxylase in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 88, 5597–5601.
- Chang, M.C.Y., Eachus, R.A., Trieu, W., Ro, D.-K., Keasling, J.D., 2007. Engineering *Escherichia coli* for production of functionalized terpenoids using plant P450s. *Nat. Chem. Biol.* 3, 274–277. doi:10.1038/nchembio875
- Chau, M., Jennewein, S., Walker, K., Croteau, R., 2004. Taxol biosynthesis: Molecular cloning and of a cytochrome p450 characterization taxoid 7 β -hydroxylase. *Chemistry & Biology* 11, 663–672. doi:10.1016/j.chembiol.2004.02.025
- Chen, C.D., Doray, B., Kemper, B., 1998. A conserved proline-rich sequence between the N-terminal signal-anchor and catalytic domains is required for assembly of functional cytochrome P450 2C2. *Arch. Biochem. Biophys.* 350, 233–238. doi:10.1006/abbi.1997.0524
- Costa, S., Almeida, A., Castro, A., Domingues, L., 2014. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8

- system. *Front Microbiol* 5, 63. doi:10.3389/fmicb.2014.00063
- Daley, D.O., Rapp, M., Granseth, E., Melen, K., Drew, D., Heijne, von, G., 2005. Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* 308, 1321–1323. doi:10.1126/science.1109730
- Davies, K.M., Deroules, S.C., 2014. Prospects for the use of plant cell cultures in food biotechnology. *Current Opinion in Biotechnology* 26, 133–140. doi:10.1016/j.copbio.2013.12.010
- Denisov, I.G., Shih, A.Y., Sligar, S.G., 2012. Structural differences between soluble and membrane bound cytochrome P450s. *J. Inorg. Biochem.* 108, 150–158. doi:10.1016/j.jinorgbio.2011.11.026
- Doray, B., Chen, C.D., Kemper, B., 2001. N-terminal deletions and His-tag fusions dramatically affect expression of cytochrome p450 2C2 in bacteria. *Arch. Biochem. Biophys.* 393, 143–153. doi:10.1006/abbi.2001.2473
- Drew, D., Lerch, M., Kunji, E., Slotboom, D.J., de Gier, J.W., 2006. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat Methods* 3, 303–313. doi:10.1038/nmeth0406-303
- Drew, D., Slotboom, D.-J., Friso, G., Reda, T., Genevaux, P., Rapp, M., Meindl-Beinker, N.M., Lambert, W., Lerch, M., Daley, D.O., Van Wijk, K.-J., Hirst, J., Kunji, E., De Gier, J.-W., 2005. A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Science* 14, 2011–2017. doi:10.1110/ps.051466205
- Drew, D.E., Heijne, von, G., Nordlund, P., de Gier, J.W., 2001. Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Lett.* 507, 220–224.
- Eugster, H.P., Bärtsch, S., Würigler, F.E., Sengstag, C., 1992. Functional co-expression of human oxidoreductase and cytochrome P450 1A1 in *Saccharomyces cerevisiae* results in increased EROD activity. *Biochem. Biophys. Res. Commun.* 185, 641–647.
- Gallage, N.J., Møller, B.L., 2015. Vanillin-Bioconversion and Bioengineering of the Most Popular Plant Flavor and Its De Novo Biosynthesis in the Vanilla Orchid. *Mol Plant* 8, 40–57. doi:10.1016/j.molp.2014.11.008

- Hamberger, B., Ohnishi, T., Hamberger, B., Séguin, A., Bohlmann, J., 2011. Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiol* 157, 1677–1695. doi:10.1104/pp.111.185843
- Hansen, E.H., Møller, B.L., Kock, G.R., Bünner, C.M., Kristensen, C., Jensen, O.R., Okkels, F.T., Olsen, C.E., Motawia, M.S., Hansen, J., 2009. De novo biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl Environ Microbiol* 75, 2765–2774. doi:10.1128/AEM.02681-08
- Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S.H., Heijne, von, G., 2005. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433, 377–381. doi:10.1038/nature03216
- Jensen, K., Osmani, S.A., Hamann, T., Naur, P., Møller, B.L., 2011. Homology modeling of the three membrane proteins of the dhurrin metabolon: catalytic sites, membrane surface association and protein-protein interactions. *Phytochemistry* 72, 2113–2123. doi:10.1016/j.phytochem.2011.05.001
- Kawate, T., Gouaux, E., 2006. Fluorescence-Detection Size-Exclusion Chromatography for Precrystallization Screening of Integral Membrane Proteins. *Structure* 14, 673–681. doi:10.1016/j.str.2006.01.013
- Koch, B.M., Sibbesen, O., Halkier, B.A., Svendsen, I., Møller, B.L., 1995. The primary sequence of cytochrome P450^{tyr}, the multifunctional N-hydroxylase catalyzing the conversion of L-tyrosine to p-hydroxyphenylacetaldehyde oxime in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench. *Arch. Biochem. Biophys.* 323, 177–186.
- Laursen, T., Møller, B.L., Bassard, J.-E., 2015. Plasticity of specialized metabolism as mediated by dynamic metabolons. *Trends Plant Sci.* 20, 20–32. doi:10.1016/j.tplants.2014.11.002
- Lee, C., Kang, H.J., Hjelm, A., Qureshi, A.A., Nji, E., Choudhury, H., Beis, K., De Gier, J.-W., Drew, D., 2014. MemStar: A one-shot *Escherichia coli*-based approach for high-level bacterial membrane protein production. *FEBS Lett.* 588, 3761–3769. doi:10.1016/j.febslet.2014.08.025

- Leonard, E., Koffas, M.A.G., 2007. Engineering of artificial plant cytochrome p450 enzymes for synthesis of isoflavones by *Escherichia coli*. *Appl Environ Microbiol* 73, 7246–7251. doi:10.1128/AEM.01411-07
- Miroux, B., Walker, J.E., 1996. Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* 260, 289–298. doi:10.1006/jmbi.1996.0399
- Moller, B.L., 2010. Dynamic metabolons. *Science*.
- Morant, M., Bak, S., Møller, B.L., Werck-Reichhart, D., 2003. Plant cytochromes P450: tools for pharmacology, plant protection and phytoremediation. *Current Opinion in Biotechnology* 14, 151–162.
- Møller, B.L., 2014. *Synthetic Biology*. Royal Society of Chemistry.
- Nour-Eldin, H.H., Hansen, B.G., Nørholm, M.H.H., Jensen, J.K., Halkier, B.A., 2006. Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments. *Nucleic Acids Res.* 34, e122–e122. doi:10.1093/nar/gkl635
- Nørholm, M.H.H., 2010. A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering. *BMC Biotechnol.* 10, 21. doi:10.1186/1472-6750-10-21
- Nørholm, M.H.H., Toddo, S., Virkki, M.T.I., Light, S., Heijne, von, G., Daley, D.O., 2013. Improved production of membrane proteins in *Escherichia coli* by selective codon substitutions. *FEBS Lett.* 587, 2352–2358. doi:10.1016/j.febslet.2013.05.063
- Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievens, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger, N.S., Newman, J.D., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496, 528–. doi:10.1038/nature12051

- Pritchard, M.P., Ossetian, R., Li, D.N., Henderson, C.J., Burchell, B., Wolf, C.R., Friedberg, T., 1997. A general strategy for the expression of recombinant human cytochrome P450s in *Escherichia coli* using bacterial signal peptides: expression of CYP3A4, CYP2A6, and CYP2E1. *Arch. Biochem. Biophys.* 345, 342–354. doi:10.1006/abbi.1997.0265
- Qi, X., Bakht, S., Qin, B., Leggett, M., Hemmings, A., Mellon, F., Eagles, J., Werck-Reichhart, D., Schaller, H., Lesot, A., Melton, R., Osbourn, A., 2006. A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18848–18853. doi:10.1073/pnas.0607849103
- Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., Las Heras, de, A., Páez-Espino, A.D., Durante-Rodríguez, G., Kim, J., Nikel, P.I., Platero, R., De Lorenzo, V., 2013. The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res.* 41, D666–75. doi:10.1093/nar/gks1119
- Sonnhammer, E.L., Heijne, von, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6, 175–182.
- Sonoda, Y., Newstead, S., Hu, N.-J., Alguel, Y., Nji, E., Beis, K., Yashiro, S., Lee, C., Leung, J., Cameron, A.D., Byrne, B., Iwata, S., Drew, D., 2011. Benchmarking Membrane Protein Detergent Stability for Improving Throughput of High-Resolution X-ray Structures. *Structure* 19, 17–25. doi:10.1016/j.str.2010.12.001
- Studier, F.W., 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* 219, 37–44.
- Studier, F.W., 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* 41, 207–234. doi:10.1016/j.pep.2005.01.016
- Sudhamsu, J., Kabir, M., Airola, M.V., Patel, B.A., Yeh, S.-R., Rousseau, D.L., Crane, B.R., 2010. Co-expression of ferrochelatase allows for complete heme incorporation into recombinant proteins produced in *E. coli*. *Protein Expression*

- and Purification 73, 78–82. doi:10.1016/j.pep.2010.03.010
- Wadsäter, M., Laursen, T., Singha, A., Hatzakis, N.S., Stamou, D., Barker, R., Mortensen, K., Feidenhans'l, R., Møller, B.L., Cárdenas, M., 2012. Monitoring shifts in the conformation equilibrium of the membrane protein cytochrome P450 reductase (POR) in nanodiscs. *Journal of Biological Chemistry* 287, 34596–34603. doi:10.1074/jbc.M112.400085
- Wagner, S., Klepsch, M.M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Hogbom, M., Van Wijk, K.J., Slotboom, D.J., Persson, J.O., de Gier, J.W., 2008. Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14371–14376. doi:10.1073/pnas.0804090105
- Waldo, G.S., Standish, B.M., Berendzen, J., Terwilliger, T.C., 1999. Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* 17, 691–695. doi:10.1038/10904
- Williams, P.A., Cosme, J., Sridhar, V., Johnson, E.F., McRee, D.E., 2000. Mammalian microsomal cytochrome P450 monooxygenase: Structural adaptations for membrane binding and functional diversity. *Molecular Cell* 5, 121–131.
- Zelasko, S., Palaria, A., Das, A., 2013. Optimizations to achieve high-level expression of cytochrome P450 proteins using *Escherichia coli* expression systems. *Protein Expression and Purification* 92, 77–87. doi:10.1016/j.pep.2013.07.017
- Zhang, H., Im, S.-C., Waskell, L., 2007. Cytochrome b(5) increases the rate of product formation by cytochrome p450 2B4 and competes with cytochrome p450 reductase for a binding site on cytochrome p450 2B4. *J. Biol. Chem.* 282, 29766–29776. doi:10.1074/jbc.M703845200

4.7 Supplementary figures

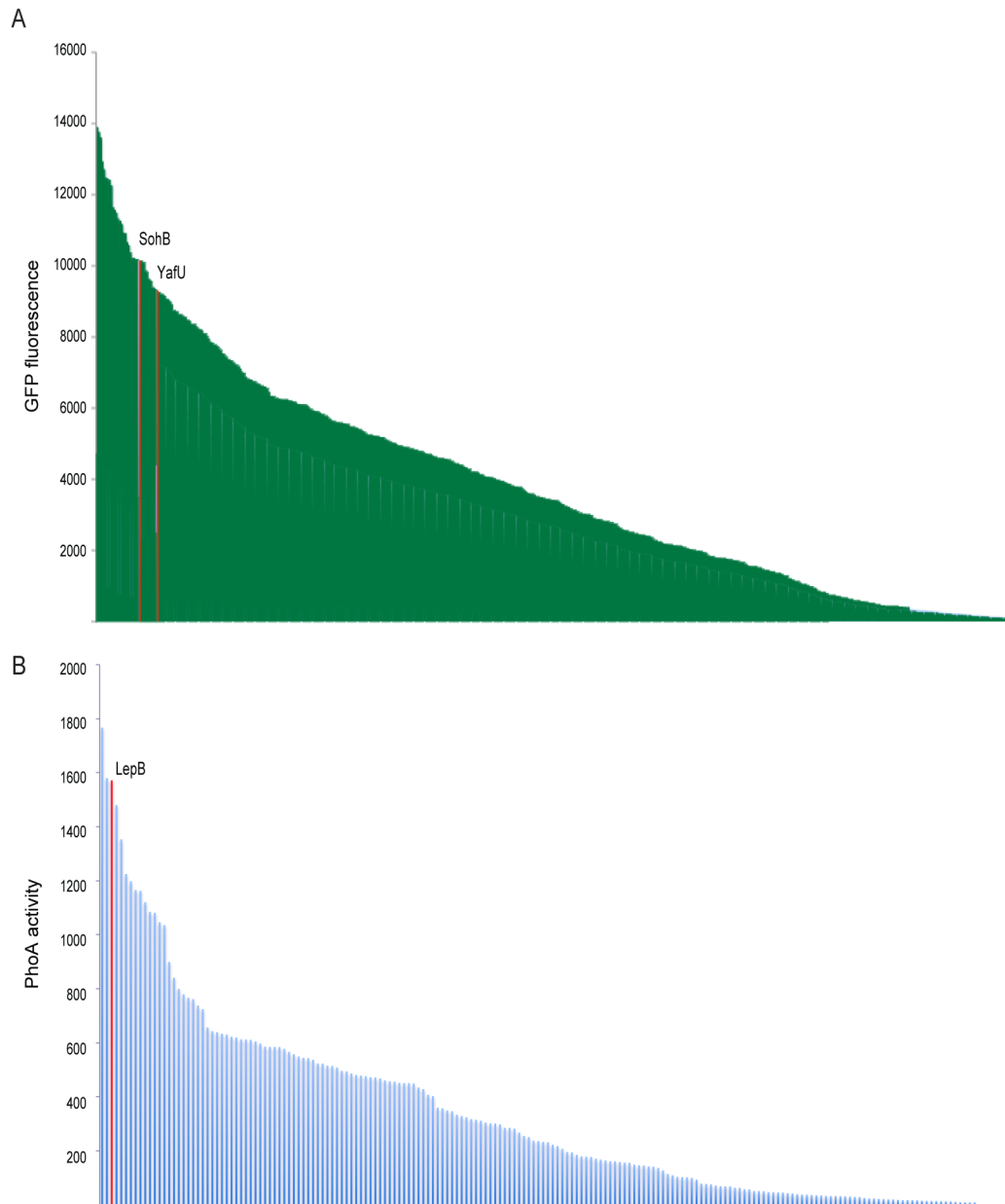


Fig. S1 Mining of a library of membrane encoding genes. (A) Production levels of 520 different membrane proteins fused to GFP, based on whole cell fluorescence data previously published {Daley:2005jg}. SohB and YafU, selected based on their high expression levels, small size and topology, are highlighted in red color. (B) Production levels of 183 different membrane proteins fused to alkaline phosphatase (PhoA), based on enzyme activity data previously published {Daley:2005jg}. LepB, selected based on the high expression level, small size and topology, is highlighted in red color.

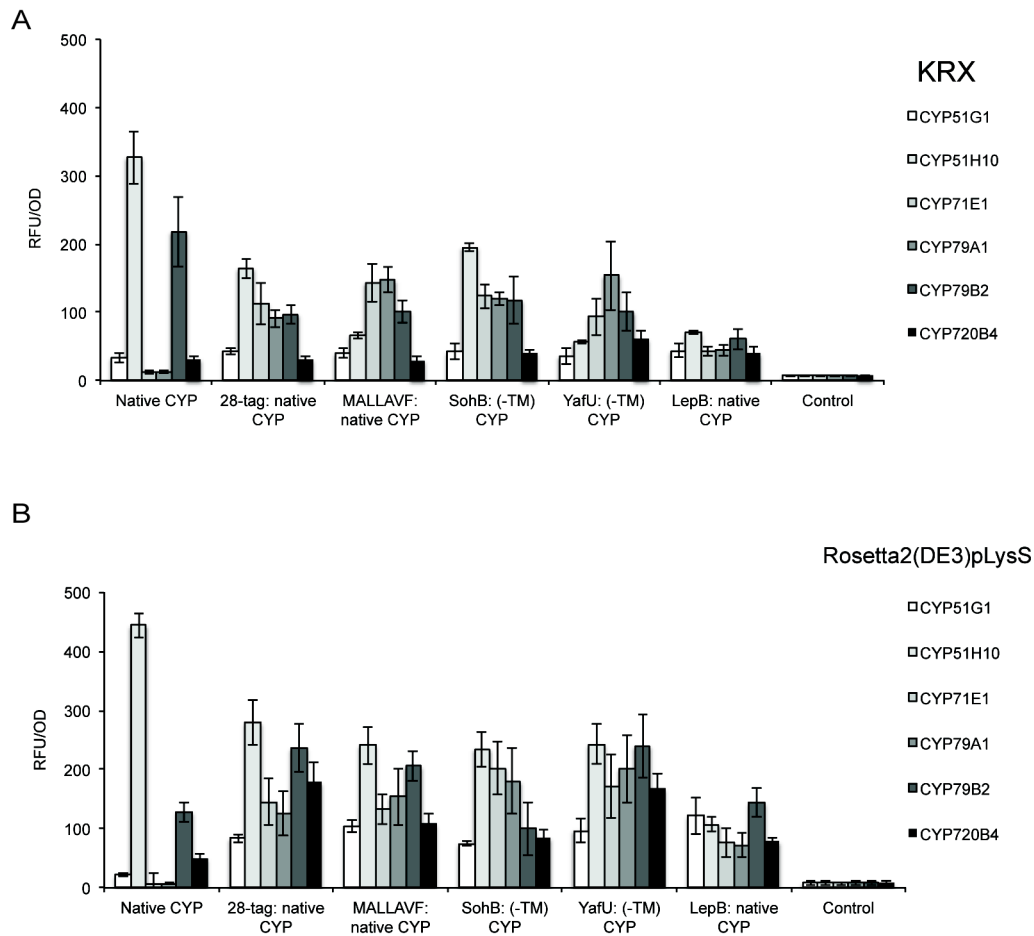


Fig. S2 Effect of different N-terminal peptides and small bacterial membrane anchors on the expression of six different *CYPs*. (A) Six *CYPs* expressed in KRX with different N-terminal modifications: (from left) native sequence as reference, 28-tag fused to native CYP sequence, Barnes-like MALLAVF peptide fused to native CYP sequence, SohB membrane domain fused to a truncated CYP sequence, YafU membrane domain fused to a truncated CYP sequence and the LepB membrane domain fused to a native sequence. Cell cultures were induced for 3 h before measuring fluorescence in a multiplate reader. (B) A parallel experiment performed in the Rosetta2(DE3)pLysS strain.

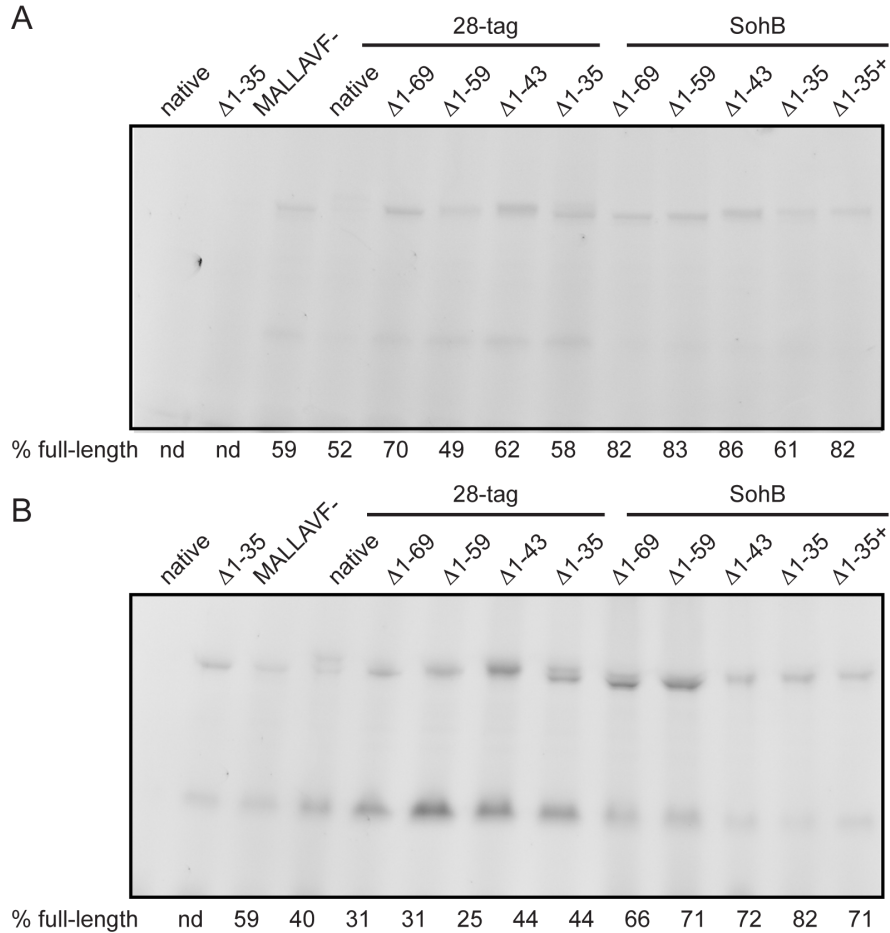


Fig. S3 Expression of differently truncated variants of CYP79A1-TEV-GFP-his8 with different N-terminal modifications. (A) 3 h induction and (B) 22 h induction of expression of native *CYP79A1*, *CYP79A1* $\Delta 1-35$, 28-tag and SohB fused to *CYP79A1*, *CYP79A1* $\Delta 1-69$, *CYP79A1* $\Delta 1-59$, *CYP79A1* $\Delta 1-43$, *CYP79A1* $\Delta 1-35$ or *CYP79A1* with an artificially extended linker region *CYP79A1* $\Delta 1-35^+$. Constructs were expressed in KRX and cell lysates analyzed by in-gel fluorescence. Quantification of full-length CYP79A1-TEV-GFP-his8 vs cleaved GFP was performed using the Image J software.

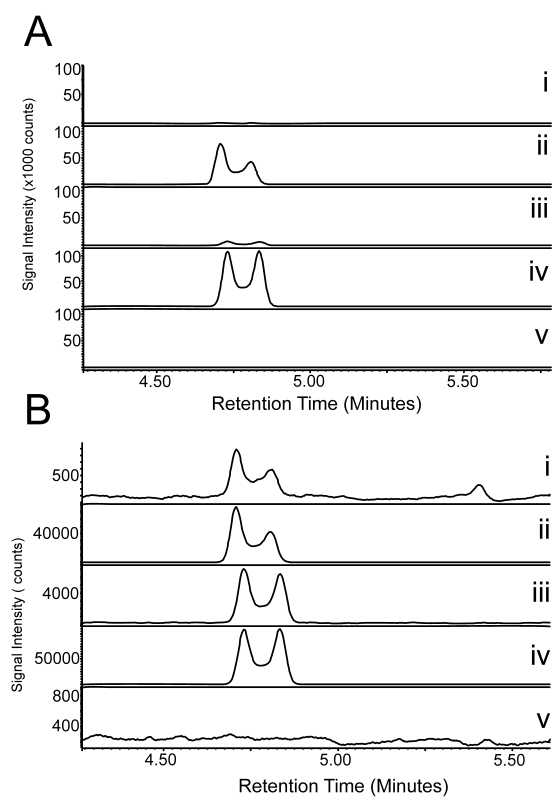


Fig. S4 High performance liquid chromatography chromatograms of i) chemical oxime standard low conc. (0.1 $\mu\text{g/mL}$), ii) chemical oxime standard high conc. (10 $\mu\text{g/mL}$), iii) sample low product conc., iv) sample high product conc. and v) blank illustrated in (A) with fixed signal intensity and (B) with independent signal intensities.

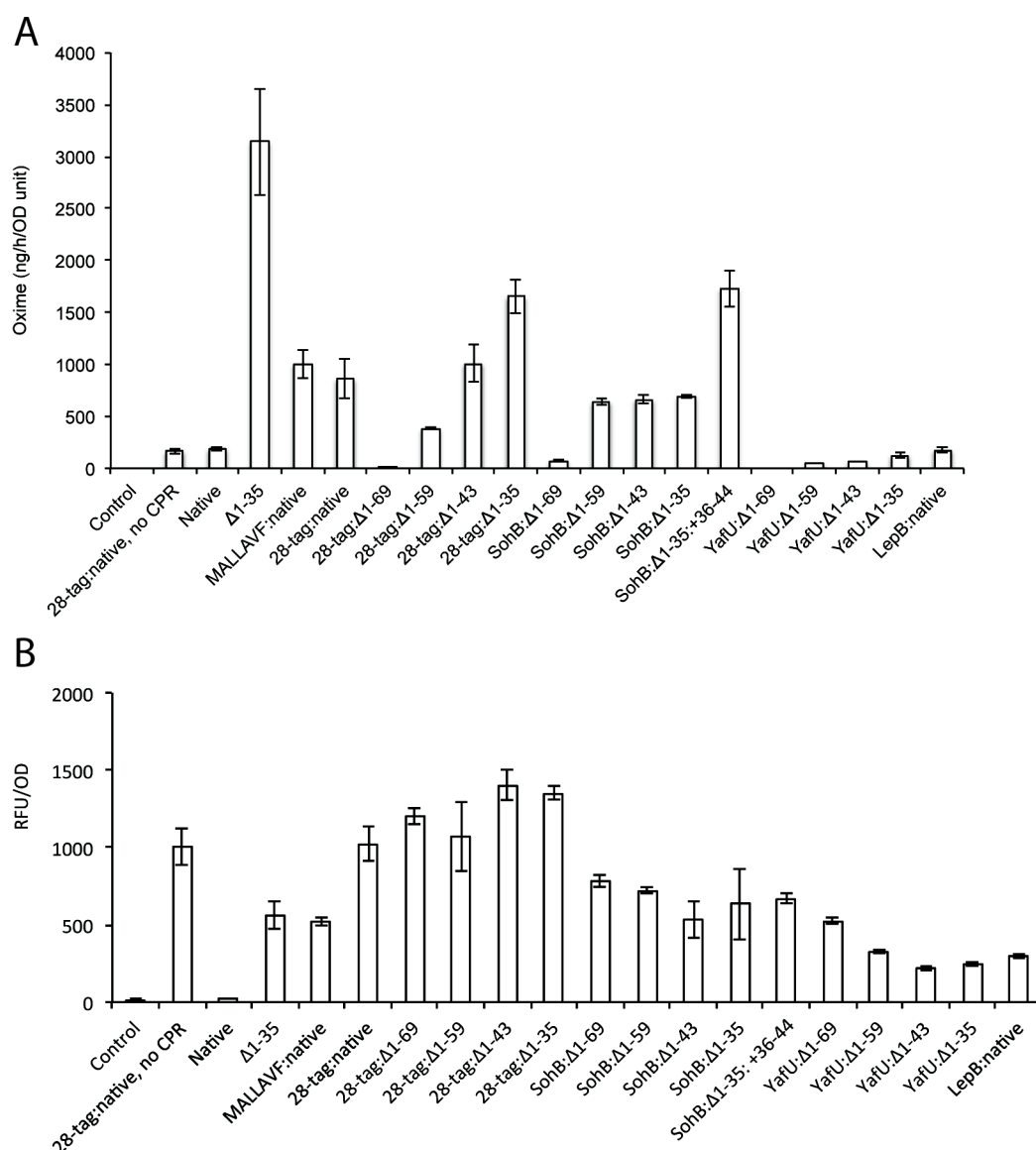


Fig. S5 Co-expression for 22 h of variants of CYP79A1-TEV-GFP-his8 with the reductase CPR2b. Constructs were expressed in KRX and cell lysates analyzed (A) *in vitro* for conversion of L-Tyrosine to (*E*)-*p*-hydroxyphenylacetaldoxime by HPLC and (B) by whole-cell fluorescence. Empty vector control and 28-tag:native CYP79A1 expressed without CPR were included as references. Error bars indicate standard error of the mean (n=3).

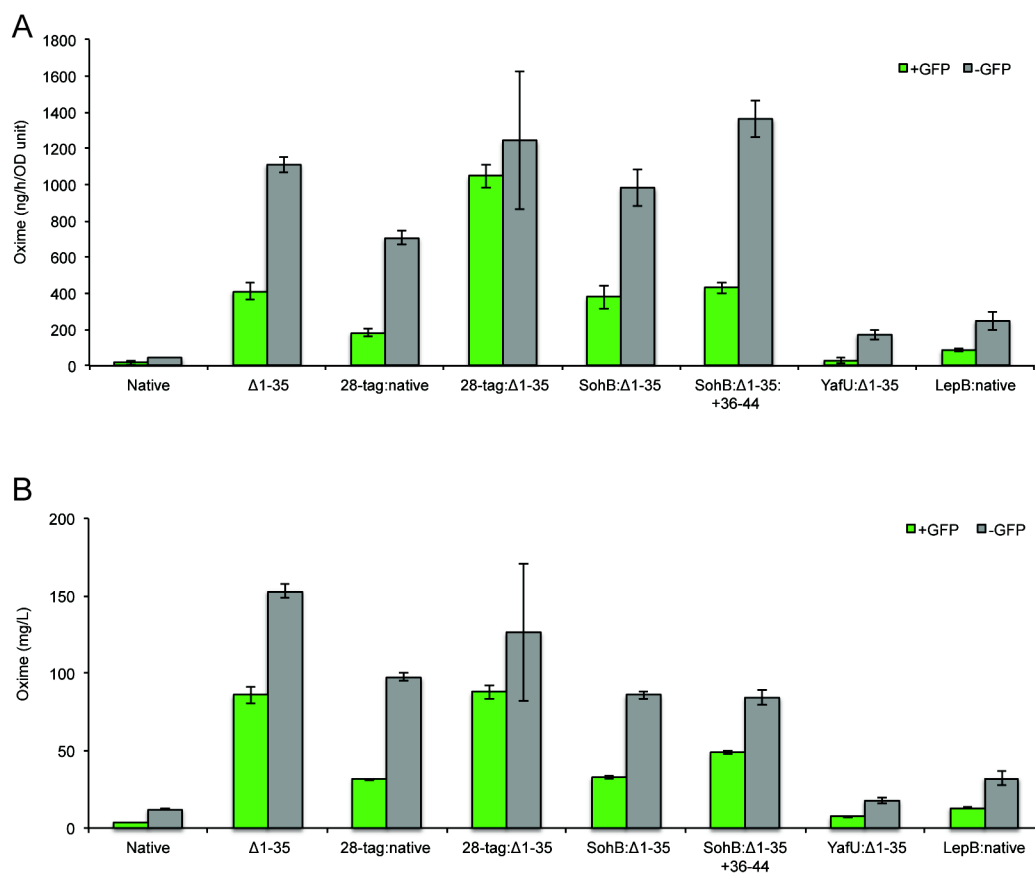


Fig. S6 Effect of the GFP moiety in the activity of different CYP79A1 variants. (A) Differently engineered versions of CYP79A1 with (green bars) or without (grey bars) fusions to GFP were co-produced with the compatible reductase electron donor CPR2b in the KRX strain and conversion of L-Tyrosine to (*E*)-*p*-hydroxyphenylacetaldoxime assayed by LC-MS. Cells were induced for 22 h, before oxime was extracted from the culture. Empty vector control and 28-tag:native CYP79A1 expressed without CPR were included as references. Error bars indicate standard error of the mean (n=3).

Chapter 4

Supplementary tables

<i>Constructs</i>	<i>Fwd oligonucleotide (5'-3')</i>	<i>Rev oligonucleotide (5'-3')</i>	<i>Template DNA</i>
<i>All pET28a(+)-based plasmids in this work had an AsiSI site removed like this</i>	ATCGCGUATTTTCGTCTCGCTC	ACGCGAUCACTGTTAA AAGGACAATTACAAA CAGGAATC	See below
<i>pET28a(+)-28tag-AsiSIcas-tev-gfp-his8</i>	ATCGCAAGCUTGCGATCGCTGCTGAGGGGGTTATATCTCCTTCTTGGA TCCAG	AGCTTGCGAUCGCCAGCTGAGGGGAGAAAACCTGTACTTCCAGGGTC	pET28a(+)-28tag <i>tev-gfp-his8</i> (Nørholm et al., 2013)
<i>pET28a(+)-MALLLLAVF-AsiSIcas-tev-gfp-his8</i>	ATCGCAAGCUTGCGATCGCTGCTGAGGGGAAA AACTGCTAATAACAG AGCCATATCATCCTCG AGTCTCCTTCTTAAAG	AGCTTGCGAUCGCCAGCTGAGGGGAGAAAACCTGTACTTCCAGGGTC	pET28a(+)-28tag- <i>AsiSIcas-tev-gfp-his8</i> (this work)
<i>pET28a(+)-sohB-AsiSIcas-tev-gfp-his8</i>	ATCGCAAGCUTGCGATCGCTGCTGAGGGGGCTG AGATTGTTGACCCGTA AC	AGCTTGCGAUCGCCAGCTGAGGGGAGAAAACCTGTACTTCCAGGGTC	pET28a(+)-sohB- <i>tev-gfp-his8</i> (Daley et al., 2005)
<i>pET28a(+)-yafU-AsiSIcas-tev-gfp-his8</i>	ATCGCAAGCUTGCGATCGCTGCTGAGGGGGCTG AGATTGTTGACCCGTA AC	AGCTTGCGAUCGCCAGCTGAGGGGAGAAAACCTGTACTTCCAGGGTC	pET28a(+)-yafU- <i>tev-gfp-his8</i> (Daley et al., 2005)
<i>pET28a(+)-tev-gfp-his8</i>	ATCCTCGAGUCTCCTTCTTAAAG	ACCTGGAUCCGAAAA CCTGTACTTC	pET28a(+)-28tag- <i>AsiSIcas-tev-gfp-his8</i> (this work)
<i>pET28a(+)-strep-HRV3C-CPR2b-tev</i>	ACCAGTGCUTCATTGACCCTGGAAGTACAGGT	AGCACTGGUCTAGCAT AACCCCTTGGGGCCTC	pET28a(+)-strep-HRV3C- CPR2b- <i>tev-gfp-his8</i>

Table S1 DNA constructs made by single fragment PCR and uracil excision

Chapter 4

Oligonucleotide name and use	Oligonucleotide sequence (5'-3')
For cloning into AsiSI cassette	
<i>CYP720B4_fwd</i>	GCAGCGAUGGCGCCCATGGCAGACCAAATAT
<i>CYP79A1_fwd</i>	GCAGCGAUGGCGACAATGGAGGTAGAGGCCG
<i>CYP79B2_fwd</i>	GCAGCGAUGAACACTTTTACCTCAAACCTCTT
<i>CYP79F1_fwd</i>	GCAGCGAUGATGAGCTTTACCACATCATTAC
<i>CYP71E1_fwd</i>	GCAGCGAUGGCCACCACCGCCACCCCGCAGC
<i>CYP720B4_rev</i>	GCTGGCGAUCCTTCATTCTCTACTCTACCATGA
<i>CYP79A1_rev</i>	GCTGGCGAUCCGATGGAGATGGACGGGTAGAGG
<i>CYP79B2_rev</i>	GCTGGCGAUCCTTCACCGTCGGGTAGAGATGC
<i>CYP79F1_rev</i>	GCTGGCGAUCCTCTCTTCTTTTCTTTCTTA
<i>CYP71E1_rev</i>	GCTGGCGAUCCGGCGGCGCGGGCGGTTCTTGAT
For cloning truncated CYPs into AsiSI cassette	
<i>tCYP720B4_fwd</i>	GCAGCGAUGCCTGGGTCTGACTGGATGGCCGC
<i>t1CYP79A1_fwd</i>	GCAGCGAUGCCACGCAAAAGCACCACCAAGT
<i>t2CYP79A1_fwd</i>	GCAGCGAUGCCGGCCGGCGTTGGCAACCCGC
<i>t3CYP79A1_fwd</i>	GCAGCGAUGTACCTGGCCCCGAGCCCTGAGGC
<i>t4CYP79A1_fwd</i>	GCAGCGAUGTACCTGGCCCCGAGCCCTGAGGCGGCCATACCTGGCCCCGAGCCCTGAGG
<i>tCYP79B2_fwd</i>	GCAGCGAUGCCGGGTCCACAGGATGGCCGA
<i>tCYP79F1_fwd</i>	GCAGCGAUGCCTGGCCCACCAGGATGGCCCA
<i>tCYP71E1_fwd</i>	GCAGCGAUGCCGGGCCCTGCGCAGCTGCCGA
For cloning into PCR	

Chapter 4

amplified pET28-variants	
<i>nCYP720B4_fwd</i>	ACTCGAGGAUGGCGGCGCCCATGGCAGACCAAATAT
<i>nCYP79A1_fwd</i>	ACTCGAGGAUGGCGGCGACAATGGAGGTAGAGGCCG
<i>nCYP79B2_fwd</i>	ACTCGAGGAUGGCGAACACTTTTACCTCAAACCTCTT
<i>nCYP79F1_fwd</i>	ACTCGAGGAUGGCGGATGAGCTTTACCACATCATTAC
<i>nCYP71E1_fwd</i>	ACTCGAGGAUGGCGGCCACCACCGCCACCCCGCAGC
<i>nCYP720B4_rev</i>	ATCCAGGUACTTCATTCTCTACTCTACCATGA
<i>nCYP79A1_rev</i>	ATCCAGGUACGATGGAGATGGACGGGTAGAGG
<i>nCYP79B2_rev</i>	ATCCAGGUACCTTCACCGTCGGGTAGAGATGC
<i>nCYP79F1_rev</i>	ATCCAGGUACTCTCTCTCTCTTTTTTTTCTTA
<i>nCYP71E1_rev</i>	ATCCAGGUACGGCGGCGCGGCGGTTCTTGTAT
<i>strepCPR2b_fwd</i>	ACTCGAGGAUGGCGGCAAGCTGGAGCCACCCGCAGT
<i>CPR2b_rev</i>	ATCCAGGUACCCAGACATCACGCAGATAGCGG

Table S2 Gene specific oligonucleotides used in this work

Chapter 5

Concluding remarks

The work presented in this thesis will have immediate and long-term impact on the way cell factories are being developed. The aim of this thesis has been to improve the premises for cell factory development by optimizing the cloning strategies, strain choice and development, as well as construct design for optimal expression of genes. The relevance of the conducted research to the field of biotechnology is covered, as well as necessary scientific background and history. The work of Chapter 2-4 describe our efforts to bring premises for cell factories to a new level. In Chapter 2, important effects of tRNA overexpression are revealed. The results offer a new perspective on a very popular expression system, and it is our ambition that the publication of this will influence how researchers consider their choice of expression system. This expanded understanding of side-effects explains the random success that we and others have experienced when applying this expression system. We can speculate that careful re-design of the pRARE plasmid might overcome some of the challenges, but more research is required to support this

In Chapter 3, a cloning strategy is described, which is designed to overcome the paradox of loss of flexibility due to standardization. We demonstrate that standardized genes can perform as well as the native versions, and the flexibility offered by this cloning strategy enables optimization of design-build-test cycles to achieve high titers of the desired compound in a cheaper and more rapid manner. Finally, Chapter 4 covers the development of a bacterial platform for the expression of the notoriously difficult cytochrome P450 monooxygenases. The platform supports engineering of complex pathways in *E. coli*, and the simple screening method can relieve the heavy workload of compound analysis for the initial steps of cell factory development.

I have conducted my research in Morten H. H. Nørholms group, Membrane Synthetic Biology, and their continued work with the subjects covered in this thesis, will help pave the way for even easier cloning, better premises for strain choice and optimized heterologous gene expression.

APPENDIX

i) Joint author statement for:

Different side effects of extra tRNA supply in standard bacterial protein production scenarios

Karina Marie Søgaard, Tonja Wolff and Morten H. H. Nørholm.

(Manuscript planned for submission)

ii) Joint author statement for:

Assembly of highly standardized gene fragments for high-level production of protoporphyrins in *E. coli*

Morten T. Nielsen, Karina M. Madsen, Susanna Seppälä, Ulla Christensen, Lone Riisberg, Scott J. Harrison, Birger Lindberg Møller and Morten H. H. Nørholm.

ACS Synth. Biol., **2015**, 4 (9), pp 959–965 **DOI:** 10.1021/acssynbio.5b00033

iii) Joint author statement for:

De-bugging and maximizing plant cytochrome P450 production in *Escherichia coli* with a scalable GFP-based optimization scheme

Ulla Christensen, Dario V. Albacete, Tonja Wolff, Morten T. Nielsen, Scott James Harrison, Anders Holmgaard Hansen, Birger Lindberg Møller, Susanna Seppälä and Morten H. H. Nørholm.

(Manuscript planned for submission)

Joint author statement

If a thesis contains articles (i.e. published journal and conference articles, unpublished manuscripts, chapters etc.) made in collaboration with other researchers, a joint-author statement verifying the PhD student's contribution to each article should be made by all authors. However, if an article has more than three authors the statement may be signed by a representative sample, cf. article 12, section 4 and 5 of the Ministerial Order No. 1039 27 August 2013 about the PhD degree. We refer to the Vancouver protocol's definition of authorship.

A representative sample of authors is comprised of


- Corresponding author and/or principal/first author (defined by the PhD student), and if there are more authors:
- 1-2 authors (preferably international/non-supervisor authors)

Titel of the article	Different side effects of extra tRNA supply in standard bacterial protein production scenarios
Author(s)	Karina M. Søgaard, Tonja Wolff, Morten H. H. Nørholm
Journal/conference * if applicable	Planned for submission
Name of PhD student	Karina Marie Søgaard
Date of Birth	1986-03-14

Description of the PhD student's contribution to the abovementioned article

Karina Marie Søgaard performed the majority of experimental design, performance of experiments, and analysis/interpretation of results, and writing the manuscript.

Signature
of the PhD student



Date

27/11-15

Signatures of co-authors

As a co-author I state that the description given above to the best of my knowledge corresponds to the process and I have no further comments.

Date
(DD/MM/YY)

Name

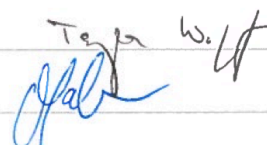
Signature

22/11-2015

TONJA WOLFF

27/11-2015

MORTEN NØRHM



Joint author statement

If a thesis contains articles (i.e. published journal and conference articles, unpublished manuscripts, chapters etc.) made in collaboration with other researchers, a joint-author statement verifying the PhD student's contribution to each article should be made by all authors. However, if an article has more than three authors the statement may be signed by a representative sample, cf. article 12, section 4 and 5 of the Ministerial Order No. 1039 27 August 2013 about the PhD degree. We refer to the Vancouver protocol's definition of authorship.

A representative sample of authors is comprised of

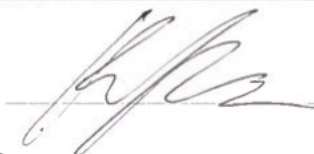
- Corresponding author and/or principal/first author (defined by the PhD student), and if there are more authors:
- 1-2 authors (preferably international/non-supervisor authors)

Titel of the article	<u>Assembly of Highly Standardized Gene Fragments for High-Level Production of Porphyrins in E. coli</u>
Author(s)	<u>Nielsen, Morten Thrane; Madsen, Karina Marie; Seppala, Susanna; Christensen, Ulla; Riisberg, Lone; Harrison, Scott James; Møller, Birger Lindberg; Nørholm, Morten</u>
Journal/conference * If applicable	<u>A C S Synthetic Biology</u>
Name of PhD student	<u>Karina Marie Søgaard</u>
Date of Birth	<u>1986-03-14</u>

Description of the PhD student's contribution to the abovementioned article

Karina Marie Søgaard contributed to experimental design, performance of experiments and analysis/interpretation of results as well as editing of the manuscript.

Signature
of the PhD student



Date

27/11-15

Signatures of co-authors

As a co-author I state that the description given above to the best of my knowledge corresponds to the process and I have no further comments.

Date (DD/MM/YY)	Name	Signature
<u>23/11/15</u>	<u>Morten Thrane Nielsen</u>	<u>Morten Thrane Nielsen</u>
<u>27/11/2015</u>	<u>Ulla J. Christensen</u>	<u>Ulla J. Christensen</u>
<u>27/11/2015</u>	<u>Morten Nørholm</u>	<u>Morten Nørholm</u>

Joint author statement

If a thesis contains articles (i.e. published journal and conference articles, unpublished manuscripts, chapters etc.) made in collaboration with other researchers, a joint-author statement verifying the PhD student's contribution to each article should be made by all authors. However, if an article has more than three authors the statement may be signed by a representative sample, cf. article 12, section 4 and 5 of the Ministerial Order No. 1039 27 August 2013 about the PhD degree. We refer to the Vancouver protocol's definition of authorship.

A representative sample of authors is comprised of

- Corresponding author and/or principal/first author (defined by the PhD student), and if there are more authors:
- 1-2 authors (preferably international/non-supervisor authors)

Titel of the article	De-bugging and maximizing plant cytochrome P450 production in <i>Escherichia coli</i> with a scalable GFP-based optimization scheme
Author(s)	Ulla Christensen, Dario V. Albacete, Karina M. Søgaard, Tonja Wolff, Morten T. Nielsen, Scott James Harrison, Anders Holmgaard Hansen, Birger Lindberg Møller, Susanna Seppälä and Morten H. H. Nørholm
Journal/conference * if applicable	Planned for submission
Name of PhD student	Karina Marie Søgaard
Date of Birth	1986-03-14

Description of the PhD student's contribution to the abovementioned article


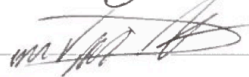
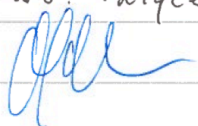
Karina Marie Søgaard contributed to experimental design, performance of experiments, and analysis/interpretation of results, as well as editing of the manuscript.

Signature
of the PhD student

 Date 27/11-15

Signatures of co-authors

As a co-author I state that the description given above to the best of my knowledge corresponds to the process and I have no further comments.

Date (DD/MM/YY)	Name	Signature
27/11-2015	Ulla J. Christensen	
27/11-2015	Dario Vazquez	
27/11-2015	 MORTEN NØRHOLM	